THE AMERICAN
MATHEMATICAL MONTHLY

THE MATHEMATICAL ASSOCIATION OF AMERICA

Volume 96, Number 1 January 1989

# Contents

(ISSN 0002-9890)

ARTICLES

Award for Distinguished Service to Ivan Niven . . . . . . . . . . . . . . KENNETH ROSS 3

Minimal Periods of Discrete and Smooth Orbits . . . . . . . . . STAVROS BUSENBERG, DAVID FISHER, AND MARIO MARTELLI 5

The Isoperimetric Inequality and Rational Approximation . . . . . . . . . . . . . . . . . . T. W. GAMELIN AND D. KHAVINSON 18

LETTERS TO THE EDITOR . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17

UNSOLVED PROBLEMS

Paradoxical Connections . . . . . . . . . . . . . . . . . . . . . ROBERT J. MACG. DAWSON 31

NOTES

On the Square Roots of Infinite Matrices . . . . . . . . . . . . LILLIAN E. PETERS HUPERT AND ANNE LEGGETT 34

A Note on Venn Diagrams . . . . . . . . . . . . . . . . . . . . . . . . . . . LEWIS PAKULA 38

Direct Sum of $J$-Rings and Zero Rings . . . . . . . . . . . . STEVE LIGH AND JIANG LUH 40

A Very Short Proof of Stirling's Formula . . . . . . . . . . . . . . . . . . . . . . . . J. M. PATIN 41

THE TEACHING OF MATHEMATICS

A Euclidean Model for Euclidean Geometry . . . . . . . . . . . . . . . . . . ADOLF MADER 43

Orthogonal Bases of $\mathbb{R}^3$ with Integer Coordinates and Integer Lengths . . . . . . . . . . . . . . . ANTHONY OSBORNE AND HANS LIEBECK 49

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 54

Advanced Problems and Solutions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 65

REVIEWS

Mathematical Problem Solving by Alan H. Schoenfeld . . . . . . . . . . . . . . . . . . . . . . . . . . STEVEN GALOVICH 68

Geometric Theory of Foliations by Cesar Camacho and Alcides Lins Neto . . . . . . . . . . . . . . ANTHONY PHILLIPS 71

TELEGRAPHIC REVIEWS . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 77

# NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Dept. of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to ANITA E. SOLOW, Department of Mathematics, Grinnell College, Grinnell, IA, 50112; in geometry or topology to DENNIS DETURCK, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems (three copies) and solutions (two copies), both elementary and advanced to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, University of Calgary, Alberta, Canada T2N 1N4.




Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: MS. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $26 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $29. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $90 per year.





Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Assocation of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

# Award for Distinguished Service to Ivan Niven

KENNETH ROSS

The 1989 Award for Distinguished Service to Mathematics is presented to Ivan Niven for his numerous contributions to American mathematics for nearly half a century.

My first exposure to Ivan was at the 1960 summer meeting at Michigan State, where he gave a wonderful set of Hedrick Lectures. He is a marvelous expositor who enjoys and loves mathematics.

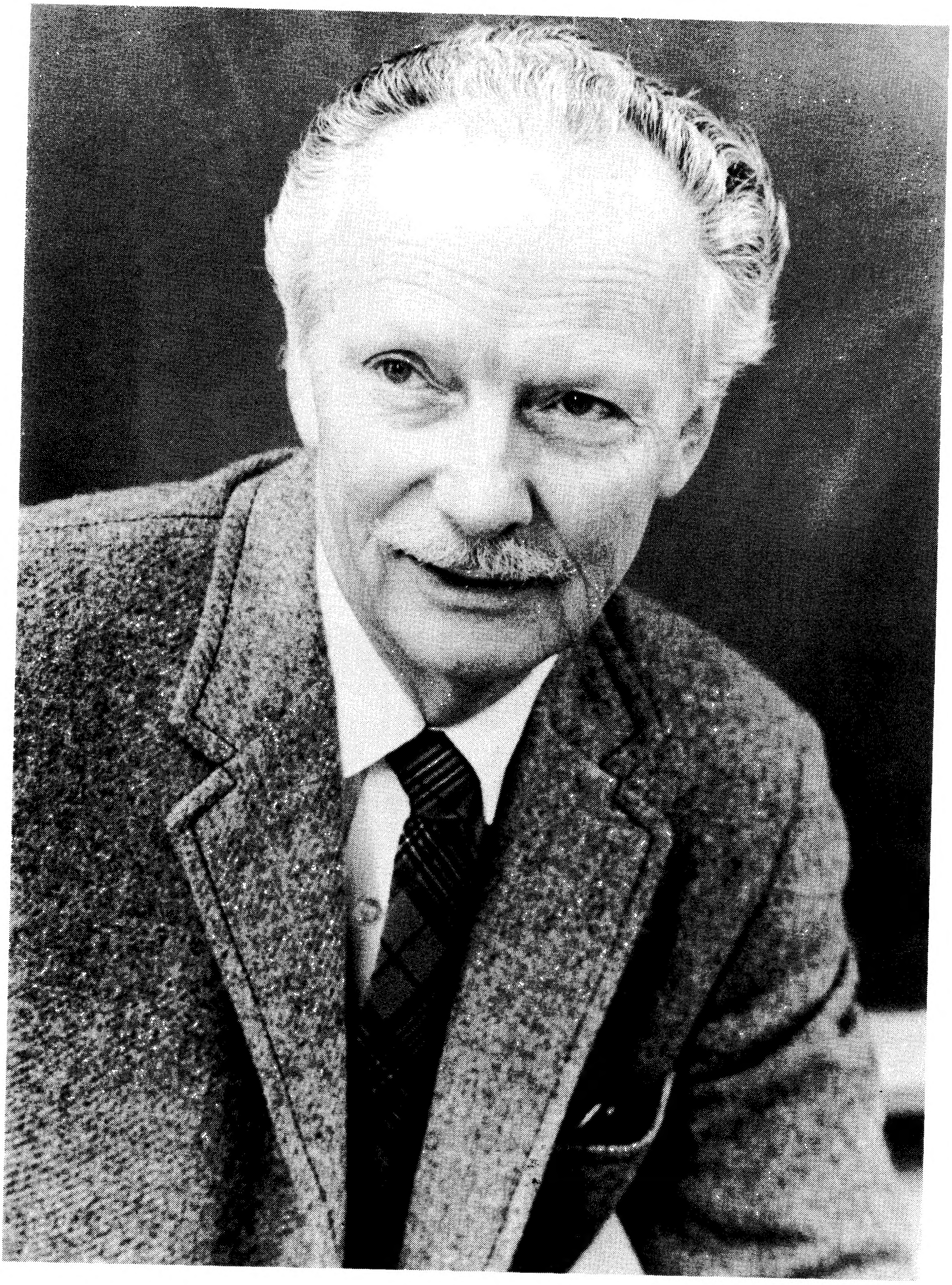
Ivan was elected First Vice-President of the Association for the years 1974–1975 and served as President in 1983–1984. He was the first President under whom I served as Secretary and he was an ideal teacher. He is patient, thoughtful, and open to new ideas. As President he was instrumental in the creation of the "Citation in honor of those who have furthered the progress of mathematics by enhancing significantly the status of women in mathematics." This special citation was the product of careful consideration in order to give well–deserved recognition while averting possible divisiveness. As President he supported the creation of the American Mathematics Project, which is taking a new approach in the improvement of the teaching of mathematics at the secondary and primary level. He was a valuable member of the Executive and Finance Committees, which serve as the "Board of Trustees" for the Board of Governors. He was a moderating influence, fiscally conservative, and cautious about new ventures.

Ivan was also active at the local level. He served as governor of the Pacific Northwest Section from 1955 to 1958 and again from 1979 to 1982. In addition, he has served on nominating committees and other ad hoc sectional committees.

Besides holding major offices in the MAA, Ivan has, over the years, served on numerous AMS and MAA committees. He was a Member-at-Large of the Council of the AMS from 1966 to 1968 and served on at least eight other AMS committees. He has been a member of the Board of Governors of the MAA since 1982 and has served on at least thirty other MAA committees. The work of these committees is essential to the governance of these organizations. Ivan brought to each of these endeavors his vigor and good judgment. His greatest contributions have probably been on committees concerned with publications of books: the New Mathematical Library series, Carus Monographs, MAA Studies in Mathematics. He has served on the Nominating Committees for editors of the American Mathematical Monthly and Mathematics Magazine. He has also assisted in selecting prizes and awards: Putnam Prize Competition, Hedrick Lecturers, this very Award for Distinguished Service to Mathematics, Lester R. Ford Awards, etc. He served on meetings committees, program committees; the list goes on and on. He also served on advisory committees to the National Science Foundation and the Office of Naval Research.

The Award for Distinguished Service to Mathematics, however, is not given for service on a maximal number of committees. It is not surprising, however, that a mathematician with Ivan's talents finds himself on so many committees.

In spite of these organizations' efforts to monopolize Ivan's time, he has a very impressive publication record. He has over sixty journal publications. Let me mention two that I've found useful: his 1947 paper containing a simple proof that $\pi$



Ivan Niven

is irrational and his 1969 *Monthly* article on formal power series. He received the Lester R. Ford Award for the latter paper. He has written chapters for two Yearbooks of the National Council of Teachers of Mathematics, in 1957 and 1987. Ivan has authored seven books, most of which have been best sellers, including the Carus Monograph *Irrational Numbers*, the Random House publications *Numbers: Rational and Irrational* and *Mathematics of Choice*, the MAA Dolciani Series publication on *Maxima and Minima Without Calculus*, and the classic text *An Introduction to the Theory of Numbers* coauthored with Herbert Zuckerman. His lean and lively 172-page book *Calculus: An Introductory Approach*, published in 1961, may well have been the book which appeared 27 years ahead of its time!

Throughout his career, Ivan has been the ultimate gentleman. In every situation he is thoughtful and wise and gives more than he receives. As must be clear from the above, he has been a "good citizen" in mathematics, in the best sense of the term.

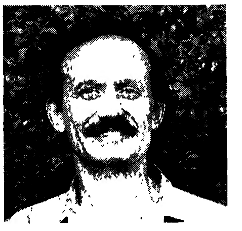
In summary, Ivan has made many significant and lasting contributions to mathematics and carried out the many tasks in an exemplary way earning him the highest respect of the entire mathematical community. Though now retired, his contributions to mathematics in many different areas continue with undiminished vigor, thoroughness, and thoughtfulness. We look forward to working with Ivan for many more years.

# Minimal Periods of Discrete and Smooth Orbits


STAVROS BUSENBERG[1] AND DAVID FISHER[2], *Harvey Mudd College, Claremont, CA*

MARIO MARTELLI[3], *Bryn Mawr College, Bryn Mawr, PA*


STAVROS BUSENBERG received engineering degrees from the Cooper Union (1962) and the Illinois Institute of Technology (1964) and his Ph.D. in Mathematics in 1967 from the latter institute under the supervision of John De Cicco. His main mathematical interests are in differential equations, nonlinear analysis, and applied mathematics.



DAVID C. FISHER received the B.S. degree in Mathematics from Harvey Mudd College in 1980. In 1985, he completed a doctorate in Applied Mathematics from the University of Maryland at College Park under the supervision of Dianne O'Leary. His research interests include parallel algorithms in numerical linear algebra, combinatorics and graph theory, and periodic solutions of differential and difference equations.

MARIO MARTELLI received his Ph.D. from the University of Florence, Italy, in 1966 under the supervision of Roberto Conti. His main mathematical interests are in differential equations, fixed point theory, and functional analysis.

**1. Continuous and discrete periodic orbits.** The uniform periodic motion of a particle on the circle $x^2 + y^2 = r^2$ is described by the system of linear differential equations

$$x'(t) = Ly(t), \qquad y'(t) = -Lx(t),$$

whose solution is $\mathbf{x}(t) = (r\sin L(t - t_0), r\cos L(t - t_0))$, for some constant $L > 0$ which represents the angular velocity. The period $T$ of this motion, that is, the minimal time required for the particle to return to its starting position, is $2\pi/L$. The


1. Partially supported by NSF Grant DMS-8703631.

2. Partially supported by NSF Grant CCR-8702553. Current Address: Department of Mathematics, University of Colorado, Denver.


3. Current Address: Department of Mathematics, California State University, Fullerton.

constant $L$ is equal to the ratio $\|\mathbf{x}'(t_1) - \mathbf{x}'(t_2)\|/\|\mathbf{x}(t_1) - \mathbf{x}(t_2)\|$ for every choice of $t_1 \neq t_2$, $0 \leqslant t_1, t_2 < T$, and therefore can be regarded as the ratio of the magnitude of the acceleration $L^2 r$ and the magnitude of the velocity $Lr$ of the particle describing this circular motion.

A natural question now arises: If a periodic motion $\mathbf{x}(t) = (x(t), y(t))$ is described by the more general system of differential equations $x' = f_1(x, y)$, $y' = f_2(x, y)$, where $f_1$ and $f_2$ are continuous with continuous partial derivatives, and if the ratio $\|\mathbf{x}'(t_1) - \mathbf{x}'(t_2)\|/\|\mathbf{x}(t_1) - \mathbf{x}(t_2)\|$ is at most $L$, then can the period be less than $2\pi/L$? The surprising negative answer to this question was given by J. Yorke in 1969 **[10]**.

THEOREM 1.1. (Yorke). *Let $E^n$ denote n-dimensional Euclidean space and let $\mathbf{f}$: $E^n \to E^n$ be continuous. Let $\mathbf{x}(t)$ be any periodic solution of period $T > 0$ of the differential equation*

$$\mathbf{x}' = \mathbf{f}(\mathbf{x}) \tag{1.1}$$

*and suppose $\mathbf{f}$ has Lipschitz constant L on the orbit described by $\mathbf{x}(t)$, that is,*

$$L = \max_{0 \leqslant t_1, t_2 \leqslant T} \frac{\|\mathbf{f}(\mathbf{x}(t_1)) - \mathbf{f}(\mathbf{x}(t_2))\|}{\|\mathbf{x}(t_1) - \mathbf{x}(t_2)\|} < \infty. \tag{1.2}$$

*Then*

$$T \geqslant 2\pi/L. \tag{1.3}$$

The norm $\|\cdot\|$ in (1.2) is the Euclidean norm $\|\mathbf{x}\| = (\langle \mathbf{x}, \mathbf{x} \rangle)^{1/2}$ obtained from the inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \Sigma_i x_i y_i$. Yorke's result, as pointed out by Lasota and Yorke **[8]** in 1971, is valid in every finite or infinite dimensional inner product space. Readers familiar with abstract inner product spaces will find that those results of this paper which are stated in $E^n$ carry over with identical proofs to inner product spaces.

The original proof of (1.3) uses some fairly sophisticated mathematical ideas. One of the aims of this paper is to give a proof based only on elementary geometry and calculus. This is done by using the discrete analog of (1.1) obtained from the "one-step Euler method" for solving (1.1) numerically. This leads to the study of the system of difference equations

$$\mathbf{x}_{i+1} = \mathbf{x}_i + h\mathbf{f}(\mathbf{x}_i) \tag{1.4}$$

and its periodic solutions, that is, those sequences $\mathbf{x}_i$ which satisfy (1.4) and are periodic of period N: $\mathbf{x}_{i+N} = \mathbf{x}_i$, $i = 1, 2, \ldots$, for some natural number $N \geqslant 1$. In order to obtain the system (1.4) from (1.1), select a time step $\Delta t = h$ and a point $\mathbf{x}_0$, as the initial position of the particle at $t = 0$, and replace $\mathbf{x}'(0)$ by $(\mathbf{x}(h) - \mathbf{x}(0))/h$ to get $\mathbf{x}_1 = \mathbf{x}(h) = \mathbf{x}_0 + h\mathbf{f}(\mathbf{x}_0)$. The difference equation (1.4) is obtained by iterating the procedure just described.

Systems of difference equations like (1.4) are important on their own right, other than as approximations to differential equations, and occur in many applications (see **[7]**, for example). The following discrete analogue of Theorem 1.1 giving a bound on the period of solutions of (1.4) was obtained in **[1]** by Busenberg and Martelli.

THEOREM 1.2. *Let* $\{\mathbf{x}_i\}_{i=0}^{\infty}$ *be a periodic solution of* (1.4) *of period* $N$ *and let* $\mathbf{f}$: $E^n \to E^n$ *have Lipschitz constant* $L$ *on the orbit* $\{\mathbf{x}_i\}$, *that is,*

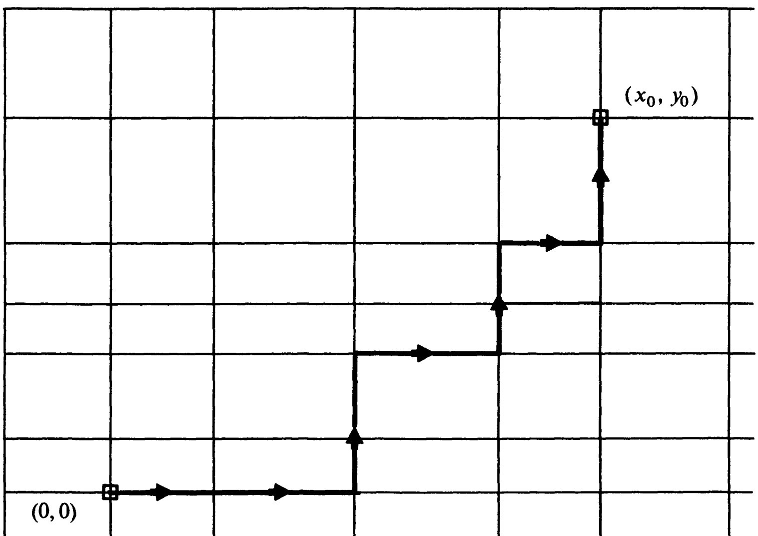
$$L = \max_{i,j} \|\mathbf{f}(\mathbf{x}_i) - \mathbf{f}(\mathbf{x}_j)\| / \|\mathbf{x}_i - \mathbf{x}_j\|. \tag{1.5}$$

*Then*

$$hL \geqslant 2\sin(\pi/N). \tag{1.6}$$

As we shall see, this theorem has a simple geometric proof. Via a limiting argument it also gives an elementary proof of its continuous analogue (1.3). The discrete result (1.6) is simpler to prove than its continuous version. Later we will encounter a situation where the opposite is true.

The Lipschitz conditions (1.2) and (1.5) use the Euclidean norm coming from the inner product of $E^n$. Some norms in $\mathbf{R}^n = \{(x_1, \dots, x_n)\colon x_i \in \mathbf{R}\}$ do not come from any inner product and it is natural to ask whether or not inequalities (1.3) and (1.6) still hold when the Lipschitz constant $L$ is obtained using one of these more general norms. For example, in $\mathbf{R}^2 = \{(x, y)\colon x, y \in \mathbf{R}\}$ one frequently considers the *taxicab norm* defined by $\|\mathbf{x}\| = \|(x, y)\| = |x| + |y|$. The name *taxicab norm* comes from the fact that, in a flat city with a rectangular grid of two-way streets, a taxi going from the origin to a street address with coordinates $(x_0, y_0)$ travels a distance $|x_0| + |y_0|$ (unless, of course, the cab driver is running up the meter).




Notice that the set of points of $\mathbf{R}^2$ at distance 1 from the origin (the unit sphere $S = \{(x, y)\colon |x| + |y| = 1\}$) differs from the familiar unit circle of the Euclidean norm, $C = \{(x, y)\colon x^2 + y^2 = 1\}$.

Recall that a norm $\|\cdot\|\colon \mathbf{E} \to [0, \infty)$, where $\mathbf{E}$ can be $\mathbf{R}^2$, $\mathbf{R}^n$, or any real linear vector space, should have the following properties: 1. $\|\mathbf{x}\| \geqslant 0$ and $\|\mathbf{x}\| = 0$ if and

only if $\mathbf{x} = \mathbf{0}$; 2. $\|t\mathbf{x}\| = |t|\,\|\mathbf{x}\|$ for all $t \in \mathbf{R}$, $\mathbf{x} \in \mathbf{E}$; 3. $\|\mathbf{x} + \mathbf{y}\| \leqslant \|\mathbf{x}\| + \|\mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y} \in \mathbf{E}$. Therefore, a normed linear space is really a pair $(\mathbf{E}, \|\cdot\|)$ where $\mathbf{E}$ is a linear vector space and $\|\cdot\|: \mathbf{E} \to [0, \infty)$ is a norm. In speaking of normed spaces, we will frequently abuse this notation and write $\mathbf{E}$ instead of the pair $(\mathbf{E}, \|\cdot\|)$.

The extension of Theorems 1.1 and 1.2 to $\mathbf{R}^n$ endowed with an arbitrary norm (or, more generally, to any normed space $\mathbf{E}$) is not without its surprises. The best possible lower bound on $T$, similar to (1.3) in Theorem 1.1, has been an open question for fifteen years. In 1971 Lasota and Yorke proved $T \geqslant 4/L$ and remarked that they did not know if this was the best possible inequality. The fact that it was not came to light when Busenberg and Martelli **[1]** proved $T \geqslant 4.5/L$. The best bound has been recently obtained by the three of us **[2]**:

THEOREM 1.3. *Let* $(\mathbf{E}, \|\cdot\|)$ *be a normed linear space and* $\mathbf{x}(t)$ *be a periodic solution of period* $T$ *of the differential equation* (1.1). *If* $\mathbf{f}$ *satisfies the Lipschitz condition* (1.2) *with respect to the above norm, then*

$$T \geqslant 6/L \tag{1.7}$$

*and* 6 *is the best possible constant in this inequality.*

The proof is surprisingly simple, requires only elementary calculus and it does not depend on a corresponding result for the difference equation (1.4). In fact, the discrete version of Theorem 1.3 is both more difficult to state and to prove, in marked contrast to the situation encountered in Theorems 1.1 and 1.2. We will state the complete result in Section 4 and present here a bound which is strict only when the period $N$ is prime.

THEOREM 1.4. *Let* $(\mathbf{E}, \|\cdot\|)$ *be a normed space and let* $\{\mathbf{x}_i\}_{i=0}^{\infty}$ *be a periodic solution of period* $N$ *of the difference equation* (1.4). *If* $\mathbf{f}$ *satisfies the Lipschitz condition* (1.5) *with respect to the above norm, then*

$$hL \geqslant \frac{12N}{N^2 - 1 + \sqrt{N^4 + 22N^2 - 23}}. \tag{1.8}$$

*This bound is the best possible when $N$ is prime.*

The bounds (1.8) for periodic solutions of periods two or three were given in **[1]** and were already known to be the best. For $N \geqslant 4$ Theorem 1.4 is new.

Section 2 contains the theorems and proofs for functions and vectors in $E^n$ (or any inner product space). The corresponding results for normed spaces are presented in sections 3 and 4. In the final section we discuss a number of open questions which can be stated in simple ways but whose proofs have eluded us so far. The reader who is tempted by these problems has our best wishes for an enjoyable and successful search for the solutions!

**2. Bounds for periods in inner product spaces.** Let $\{\mathbf{x}_i\}_{i=0}^{\infty}$ be an $N$-periodic solution of the difference equation (1.4) and let $\mathbf{v}_i = h\mathbf{f}(\mathbf{x}_{i-1}) = \mathbf{x}_i - \mathbf{x}_{i-1}$, $i = 1, 2, \dots, N$. Then

$$\begin{aligned} \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_N &= (\mathbf{x}_1 - \mathbf{x}_0) + (\mathbf{x}_2 - \mathbf{x}_1) + \cdots + (\mathbf{x}_N - \mathbf{x}_{N-1}) \\ &= \mathbf{x}_N - \mathbf{x}_0 = \mathbf{0}. \end{aligned} \tag{2.1}$$

Since $\mathbf{f}$ is Lipschitz with constant $L$ (see equation (1.5)), (1.4) implies

$$\|\mathbf{v}_{i+1} - \mathbf{v}_i\| = h\|\mathbf{f}(\mathbf{x}_i) - \mathbf{f}(\mathbf{x}_{i-1})\| \leqslant Lh\|\mathbf{x}_i - \mathbf{x}_{i-1}\|,$$

hence,

$$\|\mathbf{v}_{i+1} - \mathbf{v}_i\| \leqslant Lh\|\mathbf{v}_i\|. \tag{2.2}$$

Note, in particular, that for a solution of period two, $\mathbf{v}_1 = -\mathbf{v}_2$, hence, (2.2) gives the bound $Lh \geqslant 2$. With a little more analysis we obtain the following basic lemmas which establish Theorem 1.2.

LEMMA 2.1. *Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N$ be nonzero vectors in $E^n$ satisfying* (2.1). *Let $\theta_i \in [0, \pi]$, be the angle between $\mathbf{v}_i$ and $\mathbf{v}_{i+1}$ for $i = 1, 2, \ldots, N$, with $\mathbf{v}_{N+1} = \mathbf{v}_1$. Then*

$$\theta_1 + \theta_2 + \cdots + \theta_N \geqslant 2\pi. \tag{2.3}$$

*Proof.* The result is obvious if either $N = 2$, or if there are three consecutive vectors $\mathbf{v}_i, \mathbf{v}_{i+1}, \mathbf{v}_{i+2}$ such that $\mathbf{v}_i + \mathbf{v}_{i+1} = \mathbf{v}_{i+1} + \mathbf{v}_{i+2} = \mathbf{0}$. In any other case, assume without loss of generality that $\mathbf{v}_2 + \mathbf{v}_3 \neq \mathbf{0}$. Letting $\sphericalangle(\mathbf{x}, \mathbf{y})$ denote the angle between two nonzero vectors in $E^n$, and using the spherical inequality, we have $\sphericalangle(\mathbf{v}_1, \mathbf{v}_2 + \mathbf{v}_3) \leqslant \sphericalangle(\mathbf{v}_1, \mathbf{v}_2) + \sphericalangle(\mathbf{v}_2, \mathbf{v}_2 + \mathbf{v}_3)$ and $\sphericalangle(\mathbf{v}_2 + \mathbf{v}_3, \mathbf{v}_4) \leqslant \sphericalangle(\mathbf{v}_2 + \mathbf{v}_3, \mathbf{v}_3) + \sphericalangle(\mathbf{v}_3, \mathbf{v}_4)$. Therefore,

$$\sphericalangle(\mathbf{v}_1, \mathbf{v}_2 + \mathbf{v}_3) + \sphericalangle(\mathbf{v}_2 + \mathbf{v}_3, \mathbf{v}_4) \leqslant \theta_1 + \theta_3 + \sphericalangle(\mathbf{v}_2, \mathbf{v}_2 + \mathbf{v}_3) + \sphericalangle(\mathbf{v}_2 + \mathbf{v}_3, \mathbf{v}_3).$$

Since the vectors $\mathbf{v}_2$, $\mathbf{v}_3$ and $(\mathbf{v}_2 + \mathbf{v}_3)$ lie in one plane and $\mathbf{v}_2 + \mathbf{v}_3$ is in the interior of the angle $\sphericalangle(\mathbf{v}_1, \mathbf{v}_3)$, we get $\sphericalangle(\mathbf{v}_1, \mathbf{v}_2 + \mathbf{v}_3) + \sphericalangle(\mathbf{v}_2 + \mathbf{v}_3, \mathbf{v}_4) \leqslant \theta_1 + \theta_2 + \theta_3$. This inequality and an induction argument yield the lemma. □

LEMMA 2.2. *Let $\mathbf{v}_1, \ldots, \mathbf{v}_N$ be vectors in $E^n$ which satisfy* (2.1) *and* (2.2). *Then*

$$hL \geqslant 2\sin(\pi/N). \tag{2.4}$$

*Proof.* We start by noting that the $N$ vectors $\mathbf{w}_i \in E^{nN}$ given by

$$\mathbf{w}_1 = (\mathbf{v}_1, \ldots, \mathbf{v}_N), \qquad \mathbf{w}_2 = (\mathbf{v}_N, \mathbf{v}_1, \ldots, \mathbf{v}_{N-1}), \ldots, \qquad \mathbf{w}_N = (\mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_1)$$

satisfy the hypotheses of the lemma and also have equal norms

$$\|\mathbf{w}_i\| = \left(\sum_{i=1}^{N} \|\mathbf{v}_i\|^2\right)^{1/2} = \|\mathbf{w}_1\|, \qquad i = 1, 2, \ldots, N.$$

Hence, without loss of generality, we may assume $\|\mathbf{v}_i\| = \|\mathbf{v}_1\|$, $i = 1, 2, \ldots, N$. Since $\|\mathbf{v}_i - \mathbf{v}_{i+1}\| \leqslant Lh\|\mathbf{v}_i\| = Lh\|\mathbf{v}_1\|$, it is easy to see that $\sin(\theta_i/2) \leqslant Lh/2$. Thus, $\theta_i \leqslant 2\sin^{-1}(Lh/2)$, and (2.3) implies (2.4). □

This last result establishes Theorem 1.2. In order to prove Theorem 1.1 we select an initial point $\mathbf{x}_0$ along the orbit of a periodic solution $\mathbf{x}(t)$ to (1.1) with period $T$. We define

$$\mathbf{x}_i = \mathbf{x}_0 + \int_0^{ih} \mathbf{f}(\mathbf{x}(t))\, dt, \tag{2.5}$$

where $h = T/N$ and $N \geqslant 1$ is any natural number. Notice that the points $\mathbf{x}_i$, $i = 0, 1, \ldots, N-1$, are equally spaced in time along the orbit of $\mathbf{x}(t)$ and that $\mathbf{x}_N = \mathbf{x}_0 + \int_0^T \mathbf{f}(\mathbf{x}(t))\, dt = \mathbf{x}_0$ since $\int_0^T \mathbf{f}(\mathbf{x}(t))\, dt = \int_0^T \mathbf{x}'(t)\, dt = \mathbf{x}(T) - \mathbf{x}(0) = \mathbf{0}$. Hence, the vectors

$$\mathbf{v}_i = \mathbf{x}_i - \mathbf{x}_{i-1}, \qquad i = 1, 2, \ldots, N,$$

are all nonzero (since $T$ is the period of $\mathbf{x}(t)$) and $\mathbf{v}_1 + \cdots + \mathbf{v}_N = \mathbf{0}$. We now need to estimate $\|\mathbf{v}_{i+1} - \mathbf{v}_i\|$ so we can apply Lemma 2.2 to the vectors $\{\mathbf{v}_i\}$. Indeed, we will show

$$\|\mathbf{v}_{i+1} - \mathbf{v}_i\| \leqslant (e^{hL} - 1)\|\mathbf{v}_i\|. \tag{2.6}$$

Hence, by Lemma 2.2

$$e^{hL} - 1 \geqslant 2\sin(\pi/N) = 2\sin(h\pi/T).$$

Dividing this inequality by $h\pi/T$ and taking the limit as $h \to 0$ ($N \to \infty$) we obtain $T \geqslant 2\pi/L$, thus proving Theorem 1.1.

In order to show (2.6), we note that, by the additivity property of integrals

$$\mathbf{v}_{i+1} - \mathbf{v}_i = \int_{hi}^{h(i+1)} [\mathbf{f}(\mathbf{x}(t)) - \mathbf{f}(\mathbf{x}(t-h))]\, dt$$

and for $t \in [hi, h(i+1)]$,

$$\mathbf{x}(t) - \mathbf{x}(t-h) = \mathbf{v}_i + \int_{hi}^{t} [\mathbf{f}(\mathbf{x}(s)) - \mathbf{f}(\mathbf{x}(s-h))]\, ds.$$

Using the fact that $\mathbf{f}$ is Lipschitz continuous with constant $L$, we obtain from the first of these relations

$$\|\mathbf{v}_{i+1} - \mathbf{v}_i\| \leqslant L\int_{hi}^{h(i+1)} \|\mathbf{x}(t) - \mathbf{x}(t-h)\|\, dt \tag{2.7}$$

and from the second

$$\|\mathbf{x}(t) - \mathbf{x}(t-h)\| \leqslant \|\mathbf{v}_i\| + L\int_{hi}^{t} \|\mathbf{x}(s) - \mathbf{x}(s-h)\|\, ds. \tag{2.8}$$

Gronwall's inequality (see Lemma 2.3 below) and (2.8) imply that $\|\mathbf{x}(t) - \mathbf{x}(t-h)\| \leqslant \|\mathbf{v}_i\| e^{L(t-hi)}$, which when substituted in (2.7) yields (2.6).

For convenience we give the proof of the simplest version of the useful inequality known as Gronwall's lemma.

LEMMA 2.3. (Gronwall) *Let $g$: $[0, \infty) \to \mathbf{R}$ be continuous and suppose*

$$g(t) \leqslant b + c\int_a^t g(s)\, ds, \qquad t > a, \tag{2.9}$$

*where $b$ and $c$ are constants with $c > 0$. Then*

$$g(t) \leqslant be^{c(t-a)}, \qquad t > a. \tag{2.10}$$

*Proof.* The function $G(t) = \int_0^t g(s)\, ds$ is continuously differentiable and, by the fundamental theorem of calculus and (2.9), it satisfies

$$G'(t) - cG(t) \leqslant b. \tag{2.11}$$

Multiplying (2.11) by $e^{-ct}$ we get $(e^{-ct}G(t))' \leqslant e^{-ct}b$, which when integrated from $a$ to $t$ (note $G(a) = 0$) yields $G(t) \leqslant b(e^{c(t-a)} - 1)/c$. Substituting this inequality in (2.9) we obtain (2.10).

**3. Bounds in normed spaces.** We start by giving a simple proof of Theorem 1.3 establishing the lower bound $6/L$ for the period of periodic solutions of (1.1) in a normed space $\mathbf{E}$. The case of difference equations in $\mathbf{E}$ is quite a bit more complicated and is treated in the next section. This is contrary to the situation we encountered in $E^n$. The proof of Theorem 1.3 is based on the following lemma.

LEMMA 3.1. *Let* $\mathbf{E}$ *be a normed space and let* $\mathbf{y}:\mathbf{R}\to\mathbf{E}$ *be continuous, T-periodic and with* $\|\mathbf{y}'(t)\|$ *integrable. Then*

$$\int_0^T\int_0^T\|\mathbf{y}(t)-\mathbf{y}(s)\|\,ds\,dt \leqslant \frac{T}{6}\int_0^T\int_0^T\|\mathbf{y}'(t)-\mathbf{y}'(s)\|\,ds\,dt.$$

*Proof.* For each $s\in\mathbf{R}$, the function $g(t)=\|\mathbf{y}(t)-\mathbf{y}(s)\|$ is continuous and $T$-periodic, so its integral over a period is shift invariant:

$$\int_0^T g(t+\tau)\,dt=\int_0^T g(t)\,dt, \tag{3.1}$$

and hence

$$\begin{aligned}
A &= \int_0^T\int_0^T\|\mathbf{y}(t)-\mathbf{y}(s)\|\,ds\,dt=\int_0^T\int_0^T\|\mathbf{y}(t+s)-\mathbf{y}(s)\|\,dt\,ds\\
&= \int_0^T\int_0^T\frac{(T-t)}{T}\left\|\frac{\mathbf{y}(s+t)-\mathbf{y}(s)}{t}-\frac{\mathbf{y}(s)-\mathbf{y}(s+t-T)}{T-t}\right\|dt\,ds\\
&= \int_0^T\int_0^T\frac{(T-t)t}{T^2}\left\|\int_0^T(\mathbf{y}'(s+tr/T)-\mathbf{y}'(s+tr/T-r))\,dr\right\|ds\,dt.
\end{aligned}$$

The last step follows from the fundamental theorem of calculus. Using the triangle inequality and an interchange of order of integration we get

$$\begin{aligned}
A &\leqslant \int_0^T\int_0^T\frac{(T-t)t}{T^2}\int_0^T\|\mathbf{y}'(s+tr/T)-\mathbf{y}'(s+tr/T-r)\|\,dr\,ds\,dt\\
&= \int_0^T\frac{(T-t)t}{T^2}\int_0^T\int_0^T\|\mathbf{y}'(s+tr/T)-\mathbf{y}'(s+tr/T-r)\|\,ds\,dr\,dt.
\end{aligned}$$

Since the inner integral is over one period, it can be shifted by $\tau=tr/T-r$ without changing its value to yield:

$$\begin{aligned}
A &\leqslant \left(\int_0^T\frac{(T-t)t}{T^2}\,dt\right)\left(\int_0^T\int_0^T\|\mathbf{y}'(s+r)-\mathbf{y}'(s)\|\,ds\,dr\right)\\
&= \frac{T}{6}\int_0^T\int_0^T\|\mathbf{y}'(r)-\mathbf{y}'(s)\|\,ds\,dr.
\end{aligned}$$

This is the desired inequality. □

The proof of Theorem 1.3 proceeds as follows. For any $T$-periodic solution $\mathbf{x}(t)$ of (1.1), Lemma 3.1 and inequality (1.2) imply

$$\begin{aligned}
\int_0^T\int_0^T\|\mathbf{x}(t)-\mathbf{x}(s)\|\,ds\,dt &\leqslant \frac{T}{6}\int_0^T\int_0^T\|\mathbf{x}'(t)-\mathbf{x}'(s)\|\,ds\,dt\\
&= \frac{T}{6}\int_0^T\int_0^T\|\mathbf{f}(\mathbf{x}(t))-\mathbf{f}(\mathbf{x}(s))\|\,ds\,dt\\
&\leqslant \frac{LT}{6}\int_0^T\int_0^T\|\mathbf{x}(t)-\mathbf{x}(s)\|\,ds\,dt.
\end{aligned}$$

Hence, $T\geqslant 6/L$ and the inequality (1.7) holds.

The following example shows that 6 is the best possible constant in (1.7)

**An Example with $T = 1$ and $L = 6$:** Let $\phi$: $[0,1] \times [0,1] \to \mathbf{R}$ be defined by

$$\phi(x, y) = \begin{cases} y(1-x) & x \geqslant y \\ x(1-y) & y \geqslant x. \end{cases}$$

Observe that $\phi(0, y) = \phi(1, y) = \phi(x, 0) = \phi(x, 1) = 0$. Extend $\phi$ to a periodic function of period 1 in both variables and still call this extension $\phi$.

Let $\mathbf{E}$ be the normed space of functions $g$: $\mathbf{R}^2 \to \mathbf{R}$ which are periodic of period 1 in both variables and whose absolute value is integrable on $[0,1] \times [0,1]$, endowed with the norm

$$\|g\|_1 = \int_0^1 \int_0^1 |g(x, y)|\, dx\, dy,$$

and note that $\phi \in \mathbf{E}$. Let $\mathbf{z}$: $\mathbf{R} \to \mathbf{E}$ be the function whose value at the point $(x, y)$ is $\mathbf{z}(t)(x, y) = \phi(x + t, y)$. Note that $\mathbf{z}(t)$ is periodic of period 1 and for every $t, s \in \mathbf{R}$,

$$\|\mathbf{z}(t) - \mathbf{z}(s)\|_1 = \|\mathbf{z}(t-s) - \mathbf{z}(0)\|_1.$$

With some tedious but easy computations it can be seen that for $t \in [0,1]$

$$\|\mathbf{z}(t) - \mathbf{z}(0)\|_1 = \frac{1}{3} t(1-t).$$

Also, $\mathbf{z}$ has a continuous derivative $\mathbf{z}'$: $\mathbf{R} \to \mathbf{E}$ and

$$\|\mathbf{z}'(t) - \mathbf{z}'(0)\|_1 = 2t(1-t),$$

while

$$\|\mathbf{z}'(t) - \mathbf{z}'(s)\|_1 = \|\mathbf{z}'(t-s) - \mathbf{z}'(0)\|_1.$$

Thus we obtain

$$\|\mathbf{z}'(t) - \mathbf{z}'(s)\|_1 = 6\|\mathbf{z}(t) - \mathbf{z}(s)\|_1.$$

By setting $\mathbf{f}(\mathbf{z}(t)) = \mathbf{z}'(t)$ we have a function $\mathbf{f}$: $\Gamma \to \mathbf{E}$ defined on the trajectory $\Gamma = \{\mathbf{z}(t) \colon t \in [0,1]\}$ and which is Lipschitz with constant 6. The function $\mathbf{z}(t)$ is a periodic solution of period 1 to $\mathbf{z}'(t) = \mathbf{f}(\mathbf{z}(t))$.

The function $\mathbf{f}$ in the above example is defined and Lipschitz continuous on the orbit $\Gamma$ and, in general, it is not possible to extend a function defined on a subset of a normed space to the whole space while preserving the Lipschitz constant. In contrast, such an extension is always possible in an inner product space. However, it is possible to isometrically and isomorphically embed the space $\mathbf{E}$ in a larger normed space where the function $\mathbf{f}$ can be extended to the entire space while preserving the Lipschitz constant. So, the above example does provide us with a function defined and Lipschitz continuous on the whole normed space and for which the inequality in (1.7) reduces to an equality. This type of embedding and extension is described in Schmitt and Volkmann [**9**, pp. 404–405] where the reader can find details of how this extension can be done.

It is interesting to see why the proofs of the previous section yielding the constant $2\pi$ cannot be extended to normed spaces where the best constant is 6. The properties of angles played a crucial role in the Euclidean space proofs. In $E^n$ we define the angle $\theta$ between two nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ by setting

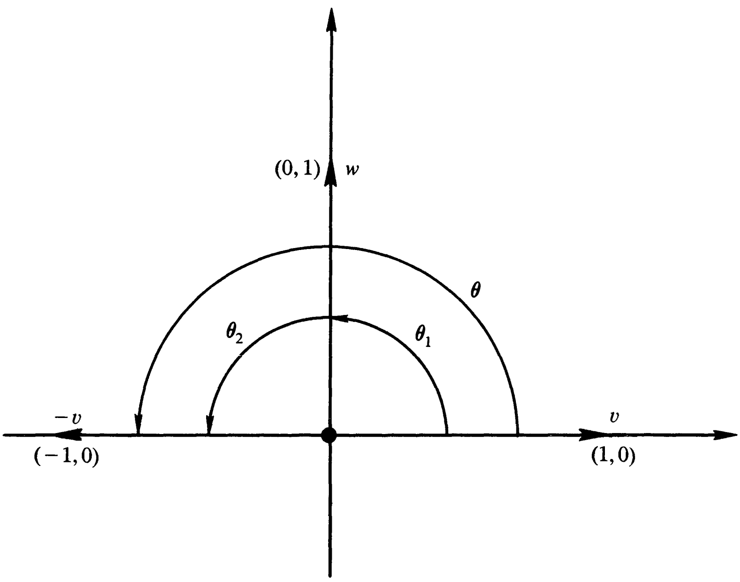
$$\cos\theta = \langle \mathbf{u}, \mathbf{v} \rangle / (\|\mathbf{u}\|\, \|\mathbf{v}\|), \tag{3.1}$$

and consider $\mathbf{u}$ and $\mathbf{v}$ to be orthogonal if, and only if, $\cos\theta = 0$. Equivalently, we could obtain $\theta$ from the less familiar formula

$$2\sin\frac{\theta}{2} = \left\| \frac{\mathbf{u}}{\|\mathbf{u}\|} - \frac{\mathbf{v}}{\|\mathbf{v}\|} \right\|, \tag{3.2}$$

and define $\mathbf{u}$ and $\mathbf{v}$ to be orthogonal if $2\sin\theta/2 = \sqrt{2}$.

This second definition of angle is valid in any normed space $\mathbf{E}$ but something unexpected happens, namely, the familiar addition formula for consecutive coplanar angles breaks down. For the vectors shown in the following figure, we see that using (3.2) and the taxicab norm, we obtain $\theta_1 = \theta_2 = \theta = \pi$. Thus $\theta_1 + \theta_2 \neq \theta$ and the proofs of the previous section cannot be extended to this case.




**4. Bounds for discrete orbits.** We now present the discrete version of Theorem 1.3. Given a positive integer $N \geqslant 2$, form the square matrix $M_N$ of size $N-1$, whose entries are

$$m_{i,j} = \min(i, N-i, ij - N\lfloor ij/N \rfloor, N - (ij - N\lfloor ij/N \rfloor)),$$

where $\lfloor x \rfloor$ denotes the maximum integer less than or equal to $x$, and hence, $ij - N\lfloor ij/N \rfloor$ is the remainder when $ij$ is divided by $N$. Notice that $m_{i,j} \geqslant 0$ and $m_{i,j} = m_{N-i,j}$, $m_{i,j} = m_{i,N-j}$. Denote by $\lambda_N$ the largest eigenvalue of $M_N$ which is known to be positive by a theorem of Perron and Frobenius. We then have the following result:

THEOREM 4.1. *Let* $\mathbf{E}$ *be a normed space*, $\mathbf{f}\colon \mathbf{E} \to \mathbf{E}$ *be Lipschitz with constant* $L$ *and let* $\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_N = \mathbf{x}_0$ *be a periodic solution of period* $N \geqslant 2$ *of the difference equation*

$$\mathbf{x}_{n+1} = \mathbf{x}_n + h\mathbf{f}(\mathbf{x}_n),$$

*where $h \geqslant 0$. Then*

$$hL \geqslant \frac{N}{\lambda_N}. \tag{4.1}$$

*Moreover, the bound in* (4.1) *is the best possible.*

The proof of inequality (4.1) uses only techniques from linear algebra and combinatorics but is rather involved and will be published elsewhere [**6**] together with some interesting properties of the eigenvalues $\lambda_N$. One of these properties is the equality

$$\lambda_N = \left(N^2 - 1 + \sqrt{N^4 + 22N^2 - 23}\right)/12 \tag{4.2}$$

which holds when $N$ is prime. So, Theorem 1.4 in the introduction is a particular case of Theorem 4.1. As can be easily seen from (4.2), and can also be proved in general, we have $N^2/\lambda_N \to 6$ as $N \to \infty$. Therefore, using the same argument as in section 2, inequality (1.7) ($TL \geqslant 6$) can also be obtained as a limiting case of (4.2) since $Nh = T$. However, the direct proof of (1.7) that we already presented is much simpler and has guided us in arriving at the discrete version (4.1) of this inequality. This is the situation mentioned in the introduction after the statement of Theorem 1.2: the discrete version of the result is much more complicated than the continuous version in this case. We have here an interesting and fruitful interplay between discrete and continuous mathematics. We established the inequality $LT \geqslant 2\pi$ as the limiting case of its easily derived discrete version $hL \geqslant 2\sin(\pi/N)$, while the inequality $TL \geqslant 6$ has helped us in the discovery of its more complicated discrete version $hL\lambda_N \geqslant N$.

The reader may be interested in seeing the form of some of the matrices $M_N$ and in knowing the behavior of the eigenvalues $\lambda_N$. For $2 \leqslant N \leqslant 5$ we obtain

$$M_2 = [1], \qquad M_3 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \qquad M_4 = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & 2 \\ 1 & 1 & 1 \end{bmatrix}, \qquad M_5 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & 1 & 2 \\ 2 & 1 & 1 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix},$$

whose characteristic polynomials are

$$p_2(\lambda) = \lambda - 1, \qquad p_3(\lambda) = \lambda^2 - 2\lambda,$$
$$p_4(\lambda) = \lambda^3 - 2\lambda^2 - 4\lambda, \qquad p_5(\lambda) = \lambda^4 - 4\lambda^3 - 4\lambda^2,$$

Using the quadratic equation we get:

$$\lambda_2 = 1, \qquad \lambda_3 = 2, \qquad \lambda_4 = 1 + \sqrt{5}, \qquad \lambda_5 = 2 + \sqrt{8}.$$

For $N \geqslant 10$ we cannot find $\lambda_N$ directly since the characteristic polynomial in this case contains an irreducible factor of degree larger than four. However, the explicit formula (4.2) for $\lambda_N$ can be obtained when $N$ is prime [3, 6], which also gives a good upper bound on all $\lambda_N$. Finally, $\lambda_N$ can be found numerically using the power method for finding the largest eigenvalue of a matrix (see Table 4.1).

Busenberg and Martelli [**1**] found that for an $N$-periodic difference equation:

$$hL \geqslant \frac{4N}{N^2 - N + 2\lfloor N/2 \rfloor}.$$

For $N = 2$ and $N = 3$, these are the same as the values given here. However, for

TABLE 4.1.

| $N$ | Inner Product Space | | Normed Space | |
|---|---|---|---|---|
| | exact | numerical | exact | numerical |
| 2 | $4$ | 4.000000 | $4$ | 4.000000 |
| 3 | $3\sqrt{3}$ | 5.196153 | $\frac{9}{2}$ | 4.500000 |
| 4 | $4\sqrt{2}$ | 5.656854 | $4(\sqrt{5}-1)$ | 4.944273 |
| 5 | $\frac{5}{2}\sqrt{10-2\sqrt{5}}$ | 5.877852 | $\frac{25}{2}(\sqrt{2}-1)$ | 5.177670 |
| 6 | $6$ | 6.000000 | ??? . | 5.409536 |
| 7 | ??? | 6.074373 | $\frac{49}{4}(\sqrt{6}-2)$ | 5.506249 |
| 8 | $8\sqrt{2-\sqrt{2}}$ | 6.122935 | ??? | 5.619933 |
| 9 | ??? | 6.156363 | ??? | 5.684398 |
| 10 | $5(\sqrt{5}-1)$ | 6.180340 | ??? | 5.741064 |
| 11 | ??? | 6.198117 | $\frac{121}{10}(\sqrt{30}-5)$ | 5.774430 |
| 12 | $12\sqrt{2-\sqrt{3}}$ | 6.211657 | ??? | 5.817793 |
| 13 | ??? | 6.222208 | $\frac{169}{14}(2\sqrt{14}-7)$ | 5.834300 |
| 14 | ??? | 6.230586 | ??? | 5.859026 |
| 15 | ??? | 6.237351 | ??? | 5.876756 |
| 16 | $16\sqrt{2-\sqrt{2-\sqrt{2}}}$ | 6.242890 | ??? | 5.891001 |
| 17 | ??? | 6.247483 | $\frac{289}{12}(\sqrt{39}-6)$ | 5.900368 |
| 18 | ??? | 6.251335 | ??? | 5.913588 |
| 19 | ??? | 6.254594 | $\frac{361}{30}(4\sqrt{15}-15)$ | 5.919598 |
| 20 | $10\sqrt{8-2\sqrt{10+2\sqrt{5}}}$ | 6.257379 | ??? | 5.929294 |
| 21 | ??? | 6.259776 | ??? | 5.934927 |
| 22 | ??? | 6.261854 | ??? | 5.940312 |
| 23 | ??? | 6.263666 | $\frac{529}{44}(\sqrt{506}-22)$ | 5.944562 |
| 24 | $24\sqrt{2-\sqrt{2+\sqrt{3}}}$ | 6.265258 | ??? | 5.950611 |
| 25 | ??? | 6.266662 | ??? | 5.953248 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| $\infty$ | $2\pi$ | 6.283186 | $6$ | 6.000000 |

$N \geqslant 4$, the present bounds are better than any previous ones and are the best possible.

Here is a table of minimum values for $NhL$ with $2 \leqslant N \leqslant 25$ for both inner product spaces and normed spaces.

In the above table, the inner product space values are found using $NhL \geqslant 2N\sin(\pi/N)$. The exact values are those which can be obtained without solving a polynomial of degree greater than two, that is, when $N$ is a power of two times either one, three or five. The numerical values for normed spaces are found by using the power method on $M_N$. When $N$ is a prime, $NhL$ can be calculated using Theorem 1.4. When $N = 4$, it can be found using the characteristic equation.

Otherwise, finding the exact value requires an exact solution of a polynomial of degree greater than two.

The reason for using $NhL$ instead of $hL$ is that $Nh$ converges to the period of a differential equation. Hence, these values converge to the minimum values for $LT$ for a differential equation.

Also, notice that for $N \geqslant 3$, the inner product space values are greater than the normed space values. This is as expected because the set of inner product norms is a subset of the set of norms. For $N = 2$, the two points forming a periodic solution lie in a one-dimensional subspace. All norms in this subspace are proportional and so the two values for $NhL$ are equal.

**5. Odds and ends.** The gap between the best constant $2\pi$ in Theorem 1.1 for inner product spaces and the best constant 6 in Theorem 1.3 for normed spaces leads to a number of questions. The first of these comes from the fact that the only examples we have been able to construct with 6 as the constant are in infinite dimensional spaces.

Question 1: Suppose that in Theorem 1.3 we add the hypothesis that $\mathbf{E}$ has finite dimension. Then, is the best constant $k$ in the inequality $LT \geqslant k$, 6 or $2\pi$ or something else?

For a particular normed space $\mathbf{E}$ the best constant $k$ in the inequality $LT \geqslant k$ is a number $k^* \geqslant 6$ (see (1.7)). Call $k^*$ the *period scale* of the space $\mathbf{E}$. Note that if $\mathbf{E}$ is a complex vector space, then the linear map $A\mathbf{x} = i\mathbf{x}$ shows that the period scale of $\mathbf{E}$ is less than or equal to $2\pi$. In fact, for any linear map $A$, the norm of $A$ equals the Lipschitz constant $L$ and, in the above case, $\|A\| = L = 1$, while $\mathbf{x}' = i\mathbf{x}$ has the periodic solution $e^{it}\mathbf{x}_0$ of period $2\pi$ for any $\mathbf{x}_0 \neq 0$ in $\mathbf{E}$. Moreover, for any linear operator $A$ on a real vector space $\mathbf{E}$, we have $\|A\|T = LT \geqslant 2\pi$. The second question asks for more details.

Question 2: Given $k > 6$ must there exist a normed space of period scale $k$? In particular, are there any spaces with $k > 2\pi$? Also, if the period scale of a normed space $\mathbf{E}$ is $2\pi$, is $\mathbf{E}$ necessarily an inner product space? What properties characterize spaces of period scale $k$?

Let us now look at the discrete results.

Question 3: In all inner product spaces with dimension at least two and for all $N$, we have an example showing that the discrete bounds of Theorem 1.2 are strict. However, given an integer $N$, the smallest dimension of a normed space for which Theorem 4.1 is known to be strict is $N - 1$. Are there normed spaces of dimension less than $N - 1$ for which equality holds in Theorem 4.1? What is the strictest inequality given $N$ and the dimension $d < N - 1$? Note that a complete answer would also answer question 1. Even the simplest nontrivial case: $N = 4$ and $d = 2$ is unresolved.

So, there still remain unknown optimal constants in this small corner of mathematics and much room for combining discrete and continuous mathematics to come up with the answers.

REFERENCES

1. S. Busenberg and M. Martelli, Bounds for the period of periodic orbits of dynamical systems, *J. Diff. Eq.*, 67 (1987) 359–71.
2. S. Busenberg, D. Fisher, and M. Martelli, Better bounds for periodic orbits of differential equations in Banach spaces, *Proc. Amer. Math. Soc.*, 86 (1986) 376–378.

3. ______, Minimal periods of discrete and smooth orbits, Harvey Mudd College Mathematics Department, *Technical Report 86-6* (1986).
4. G. D. Chakerian and M. S. Klamkin, Lengths of curves on the unit ball, *American Mathematical Monthly*, 80 (1973) 1009–1017.
5. K. Fan, O. Taussky, and J. Todd, Discrete analogs of inequalities of Wirtinger, *Mon. für Math.*, 59 (1955) 73–90.
6. D. Fisher, Minimal periods of discrete orbits, to appear, *J. Math. Anal. and Appl.*
7. H. Freeman, Discrete-Time Systems, John Wiley & Sons, Inc., New York, 1965.
8. A. Lasota and J. Yorke, Bounds for periodic solutions of differential equations in Banach spaces, *J. Diff. Eq.*, 10 (1971) 83–91.
9. K. Schmitt and P. Volkmann, Boundary value problems for second order differential equations in convex subsets of a Banach space, *Proc. Amer. Math. Soc.*, 218 (1976) 397–405.
10. J. Yorke, Periods of periodic solutions and the Lipschitz constant, *Proc. Amer. Math. Soc.*, 22 (1969) 509–512.

---

## Letters to the Editor

Editor:

I regret to bring to your attention a serious error in C. E. Van Der Ploeg's article on nonnormal quartics [*Monthly*, March 1987].

The 'theorem' on page 279 states that $K$ is a nonnormal quartic if and only if $K$ has a dihedral Galois group. However, $A_4$ and $S_4$ can appear as Galois groups for quartics, e.g., $x^4 + x + 1$.

The error in the proof occurs in the first line, where the quadratic subfield is assigned to $Q(\sqrt{t})$. Such a subfield may not exist: looking at the Galois group, we see that it occurs if and only if the subgroup of index 4 fixing $K$ is contained in a subgroup of index 2. For the possible non-abelian Galois groups of a quartic, this condition holds for the dihedral group, and not for the alternating and symmetric groups, showing the argument is circular.

The correct version of the theorem states that the following are equivalent:

(i) *$K$ is a nonnormal quartic containing a quadratic subfield*;

(ii) *$K$ has a nonnormal biquadratic minimal polynomial*;

(iii) *$K$ has a dihedral Galois group.*

As a corollary, elements of $K$ are constructible with straightedge and compass if and only if $K$ has abelian or dihedral Galois group.

Yours sincerely,

Andrew J. Lazarus

(continued on p. 67)

where the sum is over all $2^n$ of the $n$-tuples $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$. Now,

$$S_m(n) = 2^{-n} \prod_{j=1}^{n} \sum_{\varepsilon_j = \pm 1} e^{2\pi i m \varepsilon_j x} = \prod_{j=1}^{n} \frac{e^{2\pi i m x_j} + e^{-2\pi i m x_j}}{2} = \prod_{j=1}^{n} \cos(2\pi m x_j).$$

By hypothesis, the sequence $\{2mx_j\}$ has an irrational sequential limit point, say $y$. Let $I$ be a closed interval containing $y$ and not containing any integer. Then there exists a positive $\delta < 1$ such that $|\cos \pi t| < \delta$ if $t \in I$. Let $v_m(n)$ be the number of $x_i$, $1 \leqslant i \leqslant n$ such that the fractional part of $2mx_i$ lies in $I$. Then $S_m(n) \leqslant \delta^{v_m(n)} \to 0$ as $n \to \infty$.

*Editorial Comment.* All solutions received were essentially the same as Cantor's.

Solved also by Nick Martin, A. Meir (Canada), James Propp & Daniel Ullman (jointly), Arthur Rothstein, and the proposer.

---

## What Else Pythagoras Could Have Done

EDITOR:

Yoram Sagher [See reference] shows that Pythagoras could have demonstrated that the square root of an integer is an irrational or an integer without knowing anything about prime numbers. Assuming Pythagoras understood Euclid's algorithm, the following proofs show how he could have demonstrated that any integer root of an integer is an irrational or an integer, and even that the cube root of an integer either is not the root of a quadratic (i.e., not of the form $(a + b\sqrt{n})/c$) or is an integer.

In the following, all variables but $r$ are restricted to integer values.

The (shorter) proof of Sagher's special case comes first, to motivate the others. If $r = \sqrt{k} = m/n$ (in lowest terms), Euclid's algorithm gives $\alpha$ and $\beta$ for which $\alpha m + \beta n = 1$. Then $rn = m$, $rm = r^2 n = kn$, and $r = r(\alpha m + \beta n) = \alpha rm + \beta rn = \alpha kn + \beta m$, an integer.

Now take $r = m/n = \sqrt[3]{k}$, $\alpha m + \beta n = 1$. Then

$$m = rn, \qquad rn^2 = mn, \qquad rmn = m^2, \qquad rm^2 = r^3 n^2 = kn^2$$

$$r = r(\alpha m + \beta n)^2 = \alpha^2 rm^2 + 2\alpha\beta rmn + \beta^2 rn^2 = \alpha^2 kn^2 + 2\alpha\beta m^2 + \beta^2 mn,$$

an integer.

By an obvious generalization, for *any integer* $t \geqslant 2$, if $\sqrt[t]{k}$ is rational, it is an integer.

Now make the weaker assumption that $r = \sqrt[3]{k}$, and that $r$ satisfies a proper quadratic equation $ar^2 + br + c = 0$, $a \neq 0$. Then $0 = (ar^2 + br + c)(ar - b) = a^2 r^3 + (ac - b^2)r - bc$. If $ac = b^2$, divide the equation by $a^2$ to find $r^3 = (b/a)^3$ and $r = b/a$. Otherwise, put $k$ for $r^3$ and find $r = (bc - a^2 k/(ac - b^2)$. Either way, $r$ is rational, and consequently an integer.

Reference

Y. Sagher, What Pythagoras Could Have Done, *American Mathematical Monthly*, 95 (1988) 117.

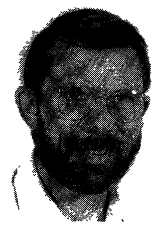
ROBERT W. FLOYD

# The Isoperimetric Inequality and Rational Approximation

T. W. GAMELIN,* *UCLA, Los Angeles, CA*

D. KHAVINSON,** *University of Arkansas, Fayetteville, AK*

T. W. GAMELIN graduated from high school in Austin, Minnesota, in 1956. After undergraduate work at Yale University (1960) and graduate work at UC Berkeley (1963), he was a Moore Instructor at MIT, and an Assistant Professor there. In 1968 he joined the Mathematics Department at UCLA, with a stint in La Plata, Argentina, in between. He has visited the Universidad Nacional de Buenos Aires, Oxford University, and the Universität des Saarlandes. His research is in function algebras.



DMITRY KHAVINSON was born in Moscow, USSR, in 1956. He received his M.S. from the Moscow State Pedagogical Institute in 1978, and his Ph.D. from Brown University in Providence, R.I. in 1983. Since 1983 he has been teaching at the University of Arkansas in Fayetteville.

**1. Introduction.** The classical "isoperimetric problem" posed by the Greeks proposes to find among all simple closed curves with a given length the curve which surrounds the largest area. The solution to this problem is contained in the *isoperimetric theorem*, which states that among all curves with a given length the circle surrounds the largest area. More precisely, let a simple closed curve with length $P$ surround a domain with area $A$. Then the circle with the same length as this curve surrounds an area equal to

$$\pi\left(\frac{P}{2\pi}\right)^2 = \frac{P^2}{4\pi}.$$

The isoperimetric theorem asserts that this area is larger than $A$, that is, that

$$P^2 \geqslant 4\pi A. \tag{1}$$

Since "cutting a hole" in the domain will increase its perimeter while diminishing its area, we conclude that (1) is actually valid for arbitrary multiply connected domains with rectifiable boundaries.

Many proofs have been given of the isoperimetric theorem, some of which are fairly brief. We wish here to give an account of a new approach, discovered in [**18**], to the isoperimetric inequality via complex Stokes' formula and a few simple ideas

*Partially supported by NSF grant #DMS-85-03780.

**Partially supported by NSF grant #DMS-84-00582.

from rational approximation theory. The approach involves associating with each compact subset $K$ of the complex plane $\mathbb{C}$ a quantity $\lambda(K)$, the *analytic content of* $K$, which expresses how well the function $\bar{z}$ can be approximated on $K$ by rational functions which are analytic on $K$. (The precise definition of the quantity $\lambda(K)$ will be presented in Section 3.)

The analytic content $\lambda(K)$ can be estimated in terms of the area $A$ of $K$ and the perimeter $P$ of $K$. We will focus on two estimates. The first is that

$$\lambda(K) \geqslant 2A/P, \tag{2}$$

which is valid when the boundary of $K$ consists of a finite number of smooth curves, or more generally whenever it is possible to make sense of the perimeter of $K$ (see [**11**, Ch. 4], [**17**]). The second is the estimate

$$\lambda(K) \leqslant \sqrt{A/\pi}, \tag{3}$$

valid for any compact set $K$, where equality holds if and only if $K$ is the union of a closed disk and a set of zero planar measure. These two estimates (2) and (3) yield immediately the isoperimetric estimate (1). From the condition for equality in (3), we then deduce the full isoperimetric theorem.

Each of the estimates (2) and (3) is interesting in its own right. In particular, the estimate (3) extends to the setting of uniform algebras, where it has been exploited by H. Alexander and others [**2**], [**3**], [**4**], [**14**], [**19**], [**30**]. In the more general context, it is closely related to Alexander's spectral area estimate.

Our program is as follows. In Section 3 we will define and discuss $\lambda(K)$. In Section 4 we will develop the complex form of Stokes' theorem and use it to establish (2). In Section 5 we will derive an estimate, due to L. Ahlfors and A. Beurling, which is of interest in other connections. In Section 6 we will establish (3), thereby completing the proof of the isoperimetric theorem. In the course of this development, the reader will see brought into play a number of important ideas from complex analysis and approximation theory, and may also find ideas for several qualifying examination problems.

Before beginning our excursion, we devote some words to placing the isoperimetric theorem in historical context.

**2. Historical background: Steiner symmetrization and Hurwitz's proof.** The Greeks already knew a version of the isoperimetric theorem and its analog in three-dimensional space. It is mentioned in the writings of Pappus (late third century A.D.), who attributes the discovery to Zenodorus. However, the first major step towards a rigorous proof was only found in the nineteenth century by J. Steiner. In 1838 Steiner [**32**], by means of an ingenious geometric method, showed that the circle solves the isoperimetric problem provided that a solution exists. Steiner purported to give several proofs of the full isoperimetric theorem, all of which stumbled over the existence of an extremal domain for the problem. In this connection, we refer to the comments of W. Blaschke [**6**, pp. 4 and 32], which close with eight quoted lines from Goethe's *Faust*.

Steiner's proofs were completed in different ways by a number of mathematicians. Of particular interest are F. Edler's completion [**10**] in 1882 of Steiner's fifth proof, and the proof given by C. Carathéodory in [**8**]. A giant step was taken by H. A. Schwarz in 1884. Inspired by work of K. Weierstrass on the existence of extrema for variational problems, Schwarz developed in a seminal paper [**29**] a

method which established rigorously the isoperimetric inequality in three-dimensional space:

$$S^3 \geqslant 36\pi V^2.$$

Here, $S$ is the surface area and $V$ is the volume of a solid. Finally, E. Schmidt **[28]** obtained the general $n$-dimensional inequality

$$S^n \geqslant 2\pi^{n/2} n^{n-1} V^{n-1} / \Gamma(n/2).$$

For more details we refer to **[5]**, **[25]**, **[26]** and the literature cited there.

To illustrate the ingenuity of geometric methods used to prove the isoperimetric inequalities even in simple situations in comparison with the apparent obviousness of the statement, we sketch a proof of (1) based on Steiner's ideas, and in particular on a process known as *Steiner symmetrization*.

Symmetrization of a plane domain $G$ with respect to a straight line $L$ changes $G$ into another domain $G^*$ characterized by the following:

(i) $G^*$ is symmetric with respect to $L$.

(ii) Any straight line orthogonal to $L$ which intersects one of the domains $G$ and $G^*$ intersects the other, and both intersections have the same length.

(iii) The intersection with $G^*$ of any straight line orthogonal to $L$ consists of just one segment bisected by $L$.

The line $L$ is called the *line of symmetrization*.

For example, a semicircle of radius $r$ symmetrized with respect to its bounding diameter is transformed into an ellipse with semiaxes $r$ and $r/2$.

Steiner observed first that symmetrization does not change the area of the domain. This is a consequence of Cavalieri's Principle, which nowadays is justified by appealing to Fubini's theorem, as follows. For purposes of argument, let us assume that $G$ is convex, and that the line of symmetrization is the $x$-axis. Then each vertical line meeting $G$ cuts the boundary in exactly two points, say $(x, y_1)$ and $(x, y_2)$, where $y_1 < y_2$. By definition the corresponding intersection with $G^*$ is a line segment whose endpoints have coordinates $(x, y)$, $(x, -y)$, where $2y = y_2 - y_1$. Therefore, according to Fubini's theorem the area $A$ of $G$ is equal to

$$A = \int_a^b (y_2 - y_1)\, dx,$$

where $a$ and $b$ are the endpoints of the interval which is the orthogonal projection of $G$ onto the $x$-axis. The domain $G^*$ is bisected by the $x$-axis, so its area $A^*$ is given by

$$A^* = 2\int_a^b y\, dx.$$

Since for each $x$, $2y = y_2 - y_1$, we have $A^* = A$.

Steiner's second fundamental observation is that symmetrization diminishes the perimeter. To see this, assume that our convex region $G$ has (say) piecewise smooth boundary with no vertical line segments. The perimeter of $G$ can then be expressed in the form

$$P = \int_a^b \left\{ \left[ 1 + \left( \frac{dy_2}{dx} \right)^2 \right]^{1/2} + \left[ 1 + \left( \frac{dy_1}{dx} \right)^2 \right]^{1/2} \right\} dx.$$

On the other hand the perimeter $P^*$ of $G^*$ is equal to

$$P^* = 2\int_a^b \left[1 + \left(\frac{dy}{dx}\right)^2\right]^{1/2} dx = \int_a^b \left[4 + \left(\frac{dy_2}{dx} - \frac{dy_1}{dx}\right)^2\right]^{1/2} dx.$$

In view of the elementary inequality

$$\left(4 + (s-t)^2\right)^{1/2} \leqslant \left(1 + s^2\right)^{1/2} + \left(1 + t^2\right)^{1/2},$$

with equality if and only if $s = -t$, the integrand in the expression for $P^*$ does not exceed the corresponding integrand for $P$, and hence $P^* \leqslant P$.

Now suppose $G$ has minimal perimeter among domains with a fixed area. Steiner symmetrization with respect to the $x$-axis then yields a domain with $P^* = P$. The condition for equality in the above inequality leads to the identity $dy_2/dx = -dy_1/dx$, so that $y_2 = -y_1 + c$. This means that $G$ is symmetric with respect to a line parallel to the $x$-axis, namely, the horizontal line which passes through the center of gravity of $G$. Symmetrizing with respect to other lines, we see that $G$ is symmetric about each line through its center of gravity, and from this it follows readily that $G$ is a circular disk. We conclude that the isoperimetric theorem is valid providing there exists a domain with reasonable boundary, whose perimeter is minimum among domains with a fixed area.

For an arbitrary convex domain $G$ with smooth boundary, we argue as follows. (For a related argument, see [**11**, Theorem 2.10.31].) Set $G_0 = G$, and let $G_1, G_2, \ldots$ be a sequence of domains, each obtained from the preceding one by Steiner symmetrization about some line through the origin, so that the perimeters of the $G_j$'s (which incidentally are convex and have smooth boundaries) converge to the minimal possible value over all choices of such sequences. It is not at all obvious but nevertheless true that the $G_j$'s converge to a smoothly bounded domain $G_\infty$, and the perimeters of the $G_j$'s decrease to that of $G_\infty$. It turns out that Steiner symmetrization does not reduce the perimeter of $G_\infty$. The earlier argument shows that $G_\infty$ is a disk, with the same area as $G$ and smaller perimeter. Since the isoperimetric inequality holds for $G_\infty$, it also holds for $G$.

This then is the idea behind Steiner's proof as completed by Carathéodory. Various other proofs, based on different geometrical ideas, can be found in [**5**], [**11**], [**26**].

It was A. Hurwitz [**15**] who supplied in 1901 the first purely analytic proof of the isoperimetric theorem. Hurwitz's proof, which features Green's formula prominently, proceeds in outline as follows. Suppose the region $G$ is bounded by a simple closed smooth curve $\Gamma$ of length $2\pi$. The idea is to parametrize $\Gamma$ by arclength, $s \to z(s)$, $0 \leqslant s \leqslant 2\pi$, and then expand the complex parameter function $z(s) = x(s) + iy(s)$ in a Fourier series

$$z(s) = \sum_{n=-\infty}^{\infty} c_n e^{ins}.$$

Since $|z'(s)| = 1$, we have

$$1 = \frac{1}{2\pi}\int_0^{2\pi} |z'(s)|^2\, ds = \sum_{n=-\infty}^{\infty} n^2 |c_n|^2.$$

On the other hand, using Green's formula we obtain

$$A = \iint_G dx\,dy = \frac{1}{2}\int_\Gamma [x\,dy - y\,dx]$$
$$= \frac{1}{2}\mathrm{Im}\int_0^{2\pi} \overline{z(s)}z'(s)\,ds = \pi \sum_{n=-\infty}^{\infty} n|c_n|^2.$$

Since $n \leqslant n^2$ for all $n$, we obtain the isoperimetric inequality $A \leqslant \pi$. Furthermore, since $n < n^2$ for all $n \neq 0, 1$, equality $A = \pi$ holds only when $z(s) = c_0 + c_1 e^{is}$ parametrizes a circle and $G$ is a disk. See **[12]** for further interesting discussion.

The first complex-analytic approach to the isoperimetric problem was developed in 1921 by T. Carleman **[9]**. For a region $G$ bounded by a smooth simple closed curve $\Gamma$, he established the inequality

$$\iint_G |f|^2\,dx\,dy \leqslant \frac{1}{4\pi}\left(\int_\Gamma |f|\,d|z|\right)^2$$

for all holomorphic functions on $G \cup \Gamma$. Applied to $f = 1$, the estimate yields the isoperimetric inequality. This estimate, valid as it is for all holomorphic functions, has the feature that it transfers from one domain to another via conformal mapping. Using the Riemann mapping theorem, one is thus obliged to verify the estimate only in the case of the unit disk. Greater detail concerning this approach and various generalizations are to be found in the Ph.D. thesis of S. Jacobs **[16]**. Carleman's estimate has been extended to finitely connected domains by S. Saitoh **[27]**, by a method which begs to be simplified.

We now turn to a quite different idea for a proof of the isoperimetric theorem.

**3. Rational approximation in $\mathbb{C}$.** Let $K$ be a compact set in $\mathbb{C}$. The algebra $R(K)$ (algebra of rational functions) consists of all rational functions with poles outside of $K$ and their uniform limits on $K$. In other words, $f \in R(K)$ if and only if there is a sequence of rational functions $p_j(z)/q_j(z)$, where $p_j$, $q_j$ are polynomials with all zeros of $q_j(z)$ outside of $K$, such that

$$\lim_{j\to\infty} \|f - p_j/q_j\|_K = 0.$$

Here $\|\cdot\|_K$ is the norm of uniform convergence on $K$,

$$\|g\|_K = \sup\{|g(z)| : z \in K\}.$$

Note that if the interior $K^o$ of $K$ is nonempty, then any function $f \in R(K)$ is analytic on $K^o$, since the approximating rational functions are analytic there. However not all continuous functions analytic on $K^o$ belong to $R(K)$, and a central problem in the theory of rational approximation is to determine exactly when a given function $f$ belongs to $R(K)$. One of the first results in this direction is the classical Runge theorem, proved by C. Runge in 1885. The version of the Runge theorem we will use is the following.

THEOREM. *If $f$ extends to be analytic in an open neighborhood of $K$, then $f \in R(K)$.*

The idea of the proof is quite simple. Any such $f$ can be represented as a complex integral

$$f(\zeta) = \frac{1}{2\pi i}\int_\Gamma \frac{f(z)}{z - \zeta}\,dz, \qquad \zeta \in K,$$

when $\Gamma$ is an appropriate contour surrounding $K$, inside of which $f$ is analytic. Since each $f(z)/(z-\zeta)$, for $z$ fixed, is a rational function of $\zeta$, one sees by writing down Riemann sums approximating the integral that $f$ is a limit of rational functions.

Another important part of Runge's theorem, which we will not use, is that if $K$ has a connected complement, then each rational function $g(\zeta) = 1/(z-\zeta)$, for $z \notin K$, can be approximated uniformly on $K$ by analytic polynomials in $\zeta$. This part of the theorem is proved by translating the pole $z$ along a path to $\infty$.

A proof of Runge's theorem based on classical approximation methods close in spirit to the original proof of Runge can be found in [**22**, Ch. 4, §2.3]. For a proof employing the functional analysis approach to approximation problems based on the idea of duality and the Hahn-Banach theorem, we refer the reader to [**13**, Ch. II, Cor. 1.2].

Besides determining when a given function $f$ belongs to $R(K)$, we might wish to be more precise and determine the distance from a given function to $R(K)$, again measured in the supremum norm on $K$:

$$\operatorname{dist}(f, R(K)) = \inf\{\|f-g\|_K : g \in R(K)\}.$$

We are particularly interested in the distance from the complex conjugate $\bar{z}$ of the coordinate function $z$ to $R(K)$; this distance we define to be the *analytic content* of $K$, denoted by $\lambda(K)$, so that

$$\lambda(K) = \operatorname{dist}(\bar{z}, R(K)) = \inf\{\|\bar{z}-g\|_K : g \in R(K)\}.$$

If $\lambda(K)=0$, then $\bar{z} \in R(K)$. Since also $z \in R(K)$, both $x = (z+\bar{z})/2$ and $y = (z-\bar{z})/2i$ belong to $R(K)$. Since $R(K)$ is an algebra, all polynomials in $x$ and $y$ are in $R(K)$. From the Stone-Weierstrass theorem we conclude that all continuous functions on $K$ are approximable by rational functions, that is, $R(K) = C(K)$. The argument is reversible, so that we obtain the following.

THEOREM. *The compact set $K$ has zero analytic content if and only if $R(K) = C(K)$.*

This observation motivates us to seek estimates for $\lambda(K)$ in terms of more geometric quantities. For instance, once we obtain the estimate (3) for $\lambda(K)$ in terms of the area of $K$, we will be able to conclude that if $K$ has zero area, then $\lambda(K) = 0$, and therefore $R(K) = C(K)$. This is a classical result of F. Hartogs and A. Rosenthal, proved in 1931. (See [**7**, Ch. III] or [**13**, Ch. II] for more detailed discussions.) Thus the estimate (3) can be regarded as a quantitative version of the Hartogs-Rosenthal theorem.

Before treating (3), we turn to the estimate (2), which will be proved using the Cauchy-Green formula.

**4. The Cauchy-Green formula.** Any student of complex analysis is familiar with Cauchy's integral representation formula. If the domain $D$ is bounded by a finite union of simple closed curves $\Gamma$, and $g$ is analytic within $D$ and across $\Gamma$, then

$$g(\zeta) = \frac{1}{2\pi i}\int_\Gamma \frac{g(z)}{z-\zeta}\,dz, \qquad \zeta \in D.$$

There is an important though less familiar version of this formula, which is valid even if $g$ is not analytic, and which includes a correction term to account for the

nonanalyticity. In terms of the differential operator

$$\frac{\partial}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right),$$

the Cauchy-Green formula is

$$g(\zeta) = \frac{1}{2\pi i}\int_\Gamma \frac{g(z)}{z-\zeta}\,dz - \frac{1}{\pi}\int\int_D \frac{\partial g}{\partial \bar{z}}\frac{1}{z-\zeta}\,dx\,dy, \qquad \zeta \in D. \tag{4}$$

The formula is valid whenever $g$ is smooth, so that we can make sense of the $\bar{z}$-derivative $\partial g/\partial \bar{z}$. Note that for fixed $\zeta$, the function $1/(z-\zeta)$ is integrable with respect to the area measure $dx\,dy$ on the bounded domain $D$, so that the area integral converges. In fact, the area integral is a convolution of the locally integrable function $1/z$ and the bounded function $\chi_D \partial g/\partial \bar{z}$, where $\chi_D$ is the characteristic function of $D$.

What happens if $g$ is analytic? By writing $g = u + iv$, where $u$ and $v$ are real, one calculates that the real and imaginary parts of the equation $\partial g/\partial \bar{z} = 0$ are precisely the Cauchy-Riemann equations for $u$ and $v$. Thus $g$ is analytic if and only if $\partial g/\partial \bar{z} = 0$. The equation $\partial g/\partial \bar{z} = 0$ can be regarded as the complex form of the Cauchy-Riemann equations. If $g$ is analytic, the correction term in (4) disappears, and (4) reduces to our old friend, the Cauchy integral formula.

For the detailed proof of (4) we refer the reader to [**7**, Ch. III, Lemma 3.1.2] or [**22**, Ch. III, §3.5]. Here we just remark that (4) can be obtained from the standard Green formula in a quite straightforward fashion, by excising from $K$ a disk $\Delta_\varepsilon$ of radius $\varepsilon$ centered at $\zeta$, applying Green's formula to the function $g(z)/(z-\zeta)$ in $K_\varepsilon = K \setminus \Delta_\varepsilon$ and then taking a limit as $\varepsilon \to 0$. This proof in particular explains why (4) is usually called the Cauchy-Green formula. We must mention here the classical papers [**23**], [**24**] of S. N. Mergelyan and [**33**] of A. G. Vitushkin, where (4) is systematically applied to various problems in approximation theory. The reader interested in possible generalizations of (4) to more general compact sets may refer to [**20**].

Now suppose we replace $g(z)$ in (4) by $(z-\zeta)g(z)$, for a fixed $\zeta \in D$. We compute using Leibniz' rule and the analyticity of $z-\zeta$ that

$$\frac{\partial}{\partial \bar{z}}[(z-\zeta)g(z)] = (z-\zeta)\frac{\partial g}{\partial \bar{z}}.$$

Hence (4) becomes

$$0 = \frac{1}{2\pi i}\int_\Gamma g(z)\,dz - \frac{1}{\pi}\int\int_D \frac{\partial g}{\partial \bar{z}}\,dx\,dy,$$

or

$$\int_\Gamma g(z)\,dz = 2i\int\int_D \frac{\partial g}{\partial \bar{z}}\,dx\,dy. \tag{5}$$

The formula (5) can be regarded as the complex form of the standard Stokes' formula, since the exterior derivative of the form $g(z)\,dz$ is given by

$$d(g(z)\,dz) = \frac{\partial g}{\partial \bar{z}}\,d\bar{z}\wedge dz = 2i\frac{\partial g}{\partial \bar{z}}\,dx\wedge dy.$$

Now we are ready to establish (2) for compact sets $K = D \cup \Gamma$ with smooth boundary $\Gamma$. Fix a function $h \in R(K)$ which is analytic in a neighborhood of $K$, and denote by $ds = d|z|$ the arclength measure on $\Gamma = \partial K$, so that the perimeter $P$ of $K$ is given by $P = \int_\Gamma ds$. Substituting in (5) the expressions

$$g(z) = \bar{z} - h(z), \qquad \frac{\partial g}{\partial \bar{z}} = 1,$$

we obtain

$$\int_\Gamma [\bar{z} - h(z)]\, dz = 2i \iint_D dx\, dy = 2iA, \tag{6}$$

where $A$ is the area of $K$. On the other hand, the obvious estimates yield

$$\left| \int_\Gamma [\bar{z} - h(z)]\, dz \right| \leqslant \int_\Gamma |\bar{z} - h(z)|\, ds$$

$$\leqslant \left[ \sup_{z \in \Gamma} |\bar{z} - h(z)| \right] \int_\Gamma ds \leqslant \|\bar{z} - h\|_K P.$$

Taking the infimum over such $h$, and taking into account (6), we obtain

$$2A \leqslant \lambda(K) P,$$

which is equivalent to (2). To complete the proof of the isoperimetric inequality, it now suffices to obtain the estimate (3).

Does the estimate (2), that $2A \leqslant \lambda(K)P$, depend on the fact that $K$ is bounded by smooth curves? Not at all. The estimate is valid for any compact set $K$, just as soon as we make sense of the perimeter $P$ of $K$, finite or infinite. For sets of finite perimeter and their geometric characterization by means of techniques from geometric measure theory, see **[11**, Ch. IV, §4 and 5]. For applications of these ideas in the context at hand, see **[17]**, **[18]**, **[20]**.

**5. The Ahlfors-Beurling estimate.** The Ahlfors-Beurling estimate is an estimate for functions of the form

$$G(\zeta) = \frac{1}{\pi} \iint_K \frac{dx\, dy}{z - \zeta}, \qquad \zeta \in \mathbb{C}, \tag{7}$$

in terms of the area of the compact set $K$. Note that the integral defining $G$ is the convolution of an $L^\infty$-function (the characteristic function of $K$) and the function $1/z$, which is locally integrable with respect to the area measure $dx\, dy$. Consequently the function $G$ depends continuously on $\zeta$. By differentiating under the integral sign, one sees that $G$ is analytic on the complement of $K$. Furthermore, $G(\zeta) \to 0$ as $\zeta \to \infty$, so that $G$ is bounded. (Supply correct proofs of these statements, and you can surely pass any analysis qualifying exam.)

There is one case in which it is easy to compute explicitly the function $G$. Suppose that $K = \Delta_\rho$ is the closed disk centered at the origin of radius $\rho > 0$. Then the associated function $G_{\Delta_\rho}$ is given by

$$G_{\Delta_\rho}(\zeta) = \frac{1}{\pi} \int_0^\rho \left[ \int_0^{2\pi} \frac{1}{re^{i\theta} - \zeta}\, d\theta \right] r\, dr.$$

The inner integral can be calculated by expanding the integrand in a geometric

series. It can be also computed with the aid of the residue calculus:

$$\int_0^{2\pi} \frac{1}{re^{i\theta} - \zeta}\, d\theta = \int_{|z|=r} \frac{1}{z-\zeta}\frac{dz}{iz} = \frac{1}{i\zeta}\int_{|z|=r}\left[\frac{1}{z-\zeta} - \frac{1}{z}\right] dz$$
$$= \begin{cases} 0, & |\zeta| < r, \\ -\dfrac{2\pi}{\zeta}, & |\zeta| > r. \end{cases}$$

Here the residues at 0 and $\zeta$ cancel if $|\zeta| < r$, while only the residue at $z = 0$ enters if $|\zeta| > r$. Thus

$$G_{\Delta_\rho}(\zeta) = -\frac{2}{\zeta}\int_0^{\min(\rho, |\zeta|)} r\, dr = \frac{-[\min(\rho, |\zeta|)]^2}{\zeta}$$
$$= \begin{cases} -\bar{\zeta}, & |\zeta| \leqslant \rho, \\ -\dfrac{\rho^2}{\zeta}, & |\zeta| \geqslant \rho. \end{cases}$$

Note that the two defining formulae coincide on the circle $|\zeta| = \rho$. In this case we easily estimate $G_{\Delta_\rho}$ in terms of the area $\pi\rho^2$ of the disk $\Delta_\rho$:

$$|G_{\Delta_\rho}(\zeta)| \leqslant \rho = \left[\frac{\operatorname{Area}(\Delta_\rho)}{\pi}\right]^{1/2}, \qquad \zeta \in \mathbf{C}, \tag{8}$$

where equality holds for $|\zeta| = \rho$.

Now the Ahlfors-Beurling estimate has exactly the same form as (8):

$$|G(\zeta)| \leqslant \left[\frac{\operatorname{Area}(K)}{\pi}\right]^{1/2}, \qquad \zeta \in \mathbf{C}, \tag{9}$$

for the $G$ defined by (7) on any compact subset $K$ of $\mathbf{C}$. Before deriving this sharp estimate, let us obtain a similar sharp estimate due to S. N. Mergelyan (see **[23]**, [7, Ch. III, Lemma 3.1.1]), which sometimes can serve in lieu of (9).

Mergelyan's estimate has the form

$$\frac{1}{\pi}\iint_K \frac{dx\,dy}{|z-\zeta|} \leqslant 2\left[\frac{\operatorname{Area}(K)}{\pi}\right]^{1/2}, \qquad \zeta \in \mathbf{C}, \tag{10}$$

with equality if and only if $K$ is the union of a closed disk centered at $\zeta$ and a set of zero area. To prove the estimate, fix $\zeta \in \mathbf{C}$, and let $D_\zeta$ be a disk centered at $\zeta$ with radius $r_\zeta = (\operatorname{Area}(K)/\pi)^{1/2}$, so that $D_\zeta$ and $K$ have the same area. Since now

$$\operatorname{Area}(D_\zeta) = \operatorname{Area}(D_\zeta \setminus K) + \operatorname{Area}(K \cap D_\zeta)$$

coincides with

$$\operatorname{Area}(K) = \operatorname{Area}(K \setminus D_\zeta) + \operatorname{Area}(K \cap D_\zeta),$$

we obtain

$$\operatorname{Area}(D_\zeta \setminus K) = \operatorname{Area}(K \setminus D_\zeta). \tag{11}$$

The idea now is to move $K \setminus D_\zeta$ back inside $D_\zeta$ to cover $D_\zeta \setminus K$, where the

integrand is larger. Then the integral appearing in (10) will increase, strictly if $K \setminus D_\zeta$ has positive area, to the corresponding integral for $D_\zeta$, which can be evaluated directly. We express this idea analytically by observing that

$$\max_{K \setminus D_\zeta} \frac{1}{|z-\zeta|} \leqslant \frac{1}{r_\zeta} \leqslant \min_{D_\zeta \setminus K} \frac{1}{|z-\zeta|},$$

so that using (11) we obtain

$$\iint_{K \setminus D_\zeta} \frac{dx\,dy}{|z-\zeta|} \leqslant \iint_{D_\zeta \setminus K} \frac{dx\,dy}{|z-\zeta|}, \tag{12}$$

with strict inequality when $K \setminus D_\zeta$ has positive area. Expressing the integral in (10) as a sum of integrals over $K \cap D_\zeta$ and $K \setminus D_\zeta$, and using (12), we find that

$$\begin{aligned} \frac{1}{\pi} \iint_K \frac{dx\,dy}{|z-\zeta|} &\leqslant \frac{1}{\pi} \iint_{K \cap D_\zeta} \frac{dx\,dy}{|z-\zeta|} + \frac{1}{\pi} \iint_{D_\zeta \setminus K} \frac{dx\,dy}{|z-\zeta|} \\ &= \frac{1}{\pi} \iint_{D_\zeta} \frac{dx\,dy}{|z-\zeta|}. \end{aligned}$$

Passing to polar coordinates and assuming without loss of generality that $\zeta = 0$, we can compute this latter integral explicitly. Its value is

$$\frac{1}{\pi} \int_0^{2\pi} \int_0^{r_\zeta} dr\,d\theta = 2\left[\frac{\operatorname{Area}(D_\zeta)}{\pi}\right]^{1/2} = 2\left[\frac{\operatorname{Area}(K)}{\pi}\right]^{1/2}.$$

This establishes Mergelyan's estimate (10).

Using (10), we may estimate our function $G$ by

$$|G(\zeta)| \leqslant \frac{1}{\pi} \iint_K \frac{dx\,dy}{|z-\zeta|} \leqslant 2\left[\frac{\operatorname{Area}(K)}{\pi}\right]^{1/2}.$$

For most applications we are happy to give up the factor 2 and use this estimate for $|G|$ in terms of area in place of the Ahlfors-Beurling estimate (9). For the isoperimetric inequality, however, we cannot afford to give up anything at all! To obtain sharp estimates, we proceed as follows.

Performing a translation if necessary, we can assume without loss of generality that $G$ attains its maximum modulus at the origin. Furthermore, by performing a rotation we can assume that $G(0) \geqslant 0$. Then

$$\|G\|_K = G(0) = \iint_K \operatorname{Re}\left(\frac{1}{z}\right) dx\,dy.$$

As is readily seen, for any $c > 0$ the set $\{\operatorname{Re}(1/z) \geqslant c\}$ is a disk centered on the real axis at the point $(1/2c, 0)$, with radius $1/2c$. Then, the same argument that we used to establish (10), shows that the integral of $\operatorname{Re}(1/z)$ over sets $K$ of a fixed area is maximized when $K$, up to a set of area zero, coincides with the disk $\Delta = \{\operatorname{Re}(1/z) \geqslant c\}$ for an appropriate $c$. For this disk $\Delta$ we have established in (8) the estimate

$$G_\Delta(0) \leqslant \left[\frac{\operatorname{Area}(\Delta)}{\pi}\right]^{1/2},$$

and since $\operatorname{Area}(K) = \operatorname{Area}(\Delta)$ and $G(0) \leqslant G_\Delta(0)$ we obtain

$$G(0) \leqslant \left[\frac{\operatorname{Area}(K)}{\pi}\right]^{1/2}.$$

We have derived the Ahlfors-Beurling estimate (9), and furthermore we see that the estimate is sharp if and only if $K$ is itself a disk, up to a set of zero area.

We invite the reader to formulate an estimate in a more general context, from which both (9) and (10) can be derived as special cases.

**6. Alexander's spectral area estimate.** Now to derive the estimate (3) we proceed as follows. Let $D$ be a bounded domain containing $K$, with smooth boundary $\Gamma$. A large disk will do. Applying the Cauchy-Green formula (4) to the function $g(z) = \bar{z}$, we obtain

$$\bar{\zeta} = \frac{1}{2\pi i}\int_\Gamma \frac{\bar{z}}{z-\zeta}\,dz - \frac{1}{\pi}\iint_D \frac{1}{z-\zeta}\,dx\,dy, \qquad \zeta \in K.$$

Thus with $G$ defined by (7), we have

$$\bar{\zeta} + G(\zeta) = \frac{1}{2\pi i}\int_\Gamma \frac{\bar{z}}{z-\zeta}\,dz - \frac{1}{\pi}\iint_{D\setminus K} \frac{1}{z-\zeta}\,dx\,dy, \qquad \zeta \in K. \tag{13}$$

The integral around $\Gamma$ depends analytically on $\zeta$ for $\zeta \in D$, so that by Runge's theorem it belongs to $R(K)$. We claim that the integral $\iint_{D\setminus K} 1/(z-\zeta)\,dx\,dy$ also belongs to $R(K)$. Indeed, if $K_\delta$ is the open $\delta$-neighborhood of $K$, then for $\delta$ small,

$$\iint_{D\setminus K} \frac{1}{z-\zeta}\,dx\,dy = \iint_{D\setminus K_\delta} \frac{1}{z-\zeta}\,dx\,dy + \iint_{K_\delta\setminus K} \frac{1}{z-\zeta}\,dx\,dy.$$

Then the first integral on the right depends analytically on $\zeta$ in $K_\delta$, so by Runge's theorem it defines a function in $R(K)$. On the other hand, Mergelyan's estimate (10) shows that the second integral is estimated by

$$\left|\iint_{K_\delta\setminus K} \frac{1}{z-\zeta}\,dx\,dy\right| \leqslant 2\left[\frac{\operatorname{Area}(K_\delta\setminus K)}{\pi}\right]^{1/2}.$$

Since $\operatorname{Area}(K_\delta \setminus K) \to 0$ as $\delta \to 0$ (why?), we find that $\iint_{D\setminus K} 1/(z-\zeta)\,dx\,dy$ is uniformly approximable on $K$ by functions analytic in a neighborhood of $K$, and hence belongs to $R(K)$.

Now the identity (13) shows that

$$\bar{z} + G \in R(K).$$

This fact incidentally was already noted by Mergelyan in [23]. (See also [33].) If we combine it with the Ahlfors-Beurling estimate (9), we obtain

$$\lambda(K) \leqslant \|\bar{z} - (\bar{z} + G)\|_K = \|G\|_K \leqslant \left[\frac{\operatorname{Area}(K)}{\pi}\right]^{1/2}.$$

This proves (3), and it completes the proof of the isoperimetric inequality. Moreover, since equality holds in the Ahlfors-Beurling estimate if and only if $K$ is the union of a closed disk and a set of zero area, we have equality in the isoperimetric inequality, among domains bounded by smooth curves, only in case the domain is a disk.

What precisely is Alexander's spectral area estimate? Let us say something about it in a general way, without clarifying all the terminology. For greater detail we refer the reader to [2], [3], [4], [14], [19], [21], [30].

Let $A$ be a uniform algebra on a compact space $X$, that is, $A$ is a closed subalgebra of $C(X)$ which contains the constant functions and which separates the points of $X$. Prototypical examples of uniform algebras are algebras of bounded analytic functions on some fixed domain, where $X$ is taken to be some compactification of the domain. The spectrum $\sigma(f)$ of $f \in A$ is the compact set of complex numbers $\lambda$ such that $1/(\lambda - f)$ does not belong to $A$. If $A$ happens to be an algebra of analytic functions, then $\sigma(f)$ often coincides with the closure of the image of $f$.

The Ahlfors-Beurling estimate for $\lambda(K)$ extends immediately to give an estimate for the distance, measured in the supremum norm on $X$, from the complex conjugate $\bar{f}$ of $f \in A$ to the algebra $A$, in terms of the area of the spectrum of $f$:

$$\operatorname{dist}(\bar{f}, A) \leqslant \left[\frac{\operatorname{Area}(\sigma(f))}{\pi}\right]^{1/2}, \qquad f \in A.$$

If now $\mu$ is a probability measure on $X$ which is multiplicative on $A$, i.e., satisfies

$$\int gh\,d\mu = \left(\int g\,d\mu\right)\left(\int h\,d\mu\right), \qquad g, h \in A,$$

then as a lower bound for $\operatorname{dist}(\bar{f}, A)$ it is easy to establish the estimate

$$\operatorname{dist}(\bar{f}, A) \geqslant \left[\int \left|f - \int f\,d\mu\right|^2 d\mu\right]^{1/2}.$$

Combining these estimates, we obtain

$$\int \left|f - \int f\,d\mu\right|^2 d\mu \leqslant \frac{\operatorname{Area}(\sigma(f))}{\pi}, \qquad f \in A. \tag{14}$$

This is *Alexander's spectral area estimate*, and in fact the proof we have presented is the original proof [2], [3].

Alexander's spectral area estimate (14) has a number of applications to analytic functions. Alexander's original application was to estimate the volumes of projections of analytic varieties. To give an idea of another type of application, recall that the BMO-norm of an analytic function on the unit disk $\Delta = \{|z| < 1\}$ is comparable to the supremum of certain $L^2$-norms,

$$\|f\|_{\mathrm{BMO}(\Delta)} \sim \sup_{|\zeta|<1} \left[\int \left|f - \int f\,d\mu_\zeta\right|^2 d\mu_\zeta\right]^{1/2},$$

where $\mu_\zeta$ is the Poisson measure (multiplicative on analytic functions!) on the unit circle $\{|z| = 1\}$ for $\zeta \in \Delta$. Thus (14) leads to estimates for the BMO-norm of an analytic function in terms of the area of the image of $f$ (not counting multiplicity!):

$$\|f\|^2_{\mathrm{BMO}} \leqslant c\,\operatorname{Area}(\operatorname{Image}(f)), \quad f \text{ analytic}.$$

Estimates of this type were originally obtained by D. Stegenga [31] using other methods. The idea of obtaining the estimate from Alexander's spectral area estimate stems from Axler and Shapiro [4]. The estimate is valid for various types of BMO-spaces connected with more general domains, in one and in several complex variables.

But this is the beginning of a new tale.

REFERENCES

1. L. Ahlfors and A. Beurling, Conformal invariants and function theoretic null sets, *Acta Math.*, 83 (1950) 101–129.
2. H. Alexander, Projections of polynomial hulls, *J. Funct. Anal.*, 3 (1973) 13–19.
3. ______, On the area of the spectrum of an element of a uniform algebra, Complex Approximation, Proceedings, Quebec, July 3–8, 1978, ed. by B. Aupetit, Birkhauser, 1980, pp. 3–12.
4. S. Axler and J. Shapiro, Putnam's theorem, Alexander's spectral area estimate, and BMO, *Math. Ann.*, 271 (1985) 161–183.
5. C. Bandle, Isoperimetric Inequalities and Applications, Pitman, Boston-London-Melbourne, 1980.
6. W. Blaschke, Kreis und Kugel, Second Edition, deGruyter, Berlin, 1956.
7. A. Browder, Introduction to Function Algebras, Benjamin, New York, 1969.
8. C. Carathéodory and E. Study, Zwei Beweise des Satzes, dass der Kreis unter allen Figuren gleichen Umfanges den grössten Inhalt hat, *Math. Ann.*, 58 (1910) 133–140.
9. T. Carleman, Zur Theorie der Minimalflächen, *Math. Zeit.*, 9 (1921) 154–160.
10. F. Edler, Vervollständigung der Steinersche elementargeometrischen Beweise für den Satz, dass der Kreis grösseren Flächeninhalt besitzt als jede andere Figur gleich grossen Umfanges, Gött. Nachr., 1882, p. 73.
11. H. Federer, Geometric Measure Theory, Springer-Verlag, Heidelberg, New York, 1969.
12. B. Fuglede, Stability in the isoperimetric problem, Matematisk Institut, Copenhagen (1986), preprint.
13. T. Gamelin, Uniform Algebras, Second Edition, Chelsea Press, 1984.
14. ______, On an estimate of Axler and Shapiro, *Math. Ann.*, 272 (1985) 189–196.
15. A. Hurwitz, Sur le problème des isopérimètres, *C. R. Acad. Sci. Paris*, 132 (1901) 401–403.
16. S. Jacobs, An isoperimetric inequality for functions analytic in multiply connected domains, Ph.D. Thesis, Institut Mittag-Leffler, 1967.
17. D. Khavinson, Sets of finite perimeter, Cauchy integrals and rational approximation, Approximation Theory IV, ed. by C. K. Chui, L. L. Schumacher and J. D. Ward, Academic Press, New York, 1983, pp. 567–574.
18. ______, Annihilating measures of the algebra $R(X)$, *J. Funct. Anal.*, 58 (1984) 175–193.
19. ______, A note on Toeplitz operators, Proceedings of the NSF-CBMS regional conference "Geometry of Banach Spaces" at the University of Missouri, Columbia, June 1984, Springer Lecture Notes in Math., Vol. 1166 (1986) 89–95.
20. ______, The Cauchy-Green formula and its applications to problems in rational approximation on sets with a finite perimeter in the complex plane, *J. Funct. Anal.*, 64 (1985) 112–123.
21. D. Khavinson and D. Luecking, On an extremal problem in the theory of rational approximation, *J. Approximation Theory*, 50 (1987) 127–132.
22. A. I. Markushevich, Theory of Analytic Functions, volumes I, II and III, English Transl. ed. by R. Silverman, Prentice-Hall, Englewood Cliffs, New Jersey, 1965.
23. S. N. Mergelyan, Uniform approximation to functions of a complex variable, *Amer. Math. Soc. Translations*, 101 (1954).
24. ______, On a theorem of M. A. Lavrentiev, *Amer. Math. Soc. Translations*, Ser. 2, 86 (1953) 3–7.
25. R. Osserman, Isoperimetric inequalities, *Bull. Amer. Math. Soc.*, 84 (1978) 1182–1238.
26. G. Pólya and G. Szegö, Isoperimetric Inequalities in Mathematical Physics, Princeton Univ. Press, Princeton, N. J., 1951.
27. S. Saitoh, The Bergman norm and the Szegö norm, *Trans. Amer. Math. Soc.*, 249 (1979) 261–279.
28. E. Schmidt, Ueber das isoperimetrische Problem in Raum von $n$ Dimensionen, *Math. Z.*, 44 (1939) 689–788.
29. H. A. Schwarz, Beweis des Satzes, dass die Kugel kleinere Oberfläche besitzt, als jeder andere Körper gleichen Volumens, Gött. Nachr., 1884, pp. 1–13; also in H. A. Schwarz, Gesammelte Abhandlungen, vol. 2, Springer Verlag, Berlin, 1890, pp. 327–340.
30. C. Stanton, Counting functions and majorization for Jensen measures, preprint.
31. D. Stegenga, A geometric condition which implies BMOA, in Harmonic Analysis in Euclidean Spaces, Proc. Symp. Pure Math., Vol. 35, Part 1, Amer. Math. Soc., Providence, 1979, pp. 427–430.
32. J. Steiner, Einfache Beweise der isoperimetrischen Hauptsätze, *J. reine ang. Math.*, 18 (1838) 281–296; also in J. Steiner, Gesammelte Werke, vol. 2, Reimer, Berlin, 1882, pp. 75–91.
33. A. G. Vitushkin, Analytic capacity of sets and problems in approximation theory, *Russian Math. Surveys*, 22 (1967) 139–200.

# UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

*In this department the* MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

## Paradoxical Connections

ROBERT J. MACG. DAWSON
*Department of Mathematics, Statistics, and Computing Science, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4H8*

A topological space $A$ is called **connected** if it can't be divided into two disjoint nonempty closed subspaces. Such a dissection is called a **separation**. Equivalently, a connected space is one with no proper open-and-closed subset (or **component**). For instance, the interval $[0,1]$ with the induced topology as a subset of the real line is connected; while $[0,1/2) \cup (1/2,1]$ is not, as $[0,1/2)$ is relatively closed and open.

As connectedness is most naturally defined as a negative property ("there does not exist..."), it is often easier to show a space to be nonconnected, by exhibiting a component or a separation, than to show a space to be connected. This is one motivation (not, of course, the only one!) for the definition of path-connectedness. A topological space $A$ is **path-connected** if, for any two points $a_0, a_1 \in A$, there exists a continuous map $f$: $[0,1] \to A$ such that $f(0) = a_0$, $f(1) = a_1$. This implies connectedness, for if $A_0$ and $A_1$ form a separation of $A$, $a_0 \in A_0$, $a_1 \in A_1$, then $f^{-1}(A_0)$ and $f^{-1}(A_1)$ form a separation of $[0,1]$, which, as we have seen, cannot occur. The opposite implication is not true; Fig. 1 shows a space that is connected but not path-connected (why?)

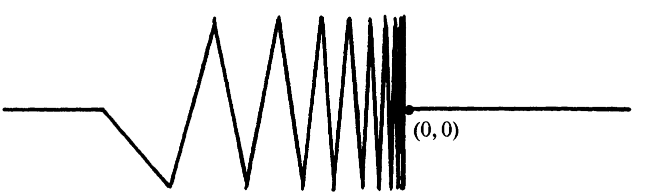



FIG. 1

If the topology of a space has a basis of connected open sets, we call it **locally connected**. This property neither implies connectedness nor is implied by it. $[0,1] \cup [2,3]$ is locally connected but not connected, while the space in Fig. 1 is not locally connected (consider a neighborhood of $(0,0)$). The "Warsaw Circle" (Fig. 2) obtained by joining opposite ends of Fig. 1 is path-connected but not locally connected. There are also spaces that are locally connected and connected but not

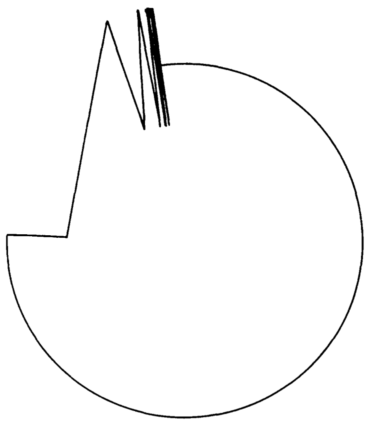


FIG. 2

path-connected; examples may be found in [**4**, p. 28], or the interested reader may enjoy constructing her own.

Another motivation for studying path-connectedness is that it corresponds well to our intuition on the behavior of connected sets. For instance, if a path-connected set $A$ in the unit square $S = [0,1] \times [0,1]$ contains $(0,0)$ and $(1,1)$ and does not contain the other two corners, it separates them, in the sense that $(0,1)$ and $(1,0)$ lie in different components of $S \setminus A$. This is proved in [**3**] by Kiang, Morrison, and Wright. However, we can find disjoint connected sets $A$ and $B$ in $S$, such that $A$ contains $(0,0)$ and $(1,1)$, while $B$ contains $(0,1)$ and $(1,0)$. We may call such a pair of disjoint connected sets **crossed**. An example is given in Appendix 2 of [**1**]; FIGURE 3 shows a slight modification of that construction, and FIGURE 4 shows a different construction with the same property. A very different construction, in which $A$ and $B$ are not only connected, but also locally connected, was given by Jones [**2**]. Unfortunately, this construction cannot be drawn, as both $A$ and $B$ are dense in the square!

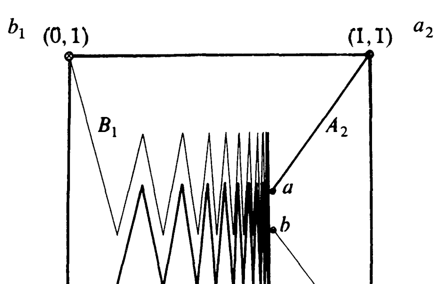



FIG. 3

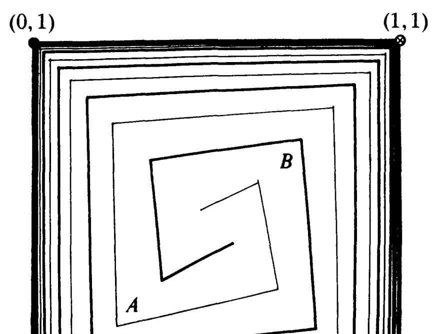



FIG. 4

A curious feature of the example in FIGURE 3 is that the union of the two sets, $A \cup B$, cannot be repartitioned into disjoint connected sets, one containing the left-hand corners $(0,0)$ and $(0,1)$, the other containing the right-hand corners $(1,0)$ and $(1,1)$. To see this, note that the removal of the points $a$ and $b$ disconnects the remainder of $A \cup B$ into 4 components $A_1$, $A_2$, $B_1$, $B_2$, and that to join $(0,0)$ to $(0,1)$ requires one of these two points to "glue" the two zigzags $A_1$ and $B_1$. On the other hand, any connected subset of $A \cup B$ containing $(1,1)$ must contain $a$ or lie entirely within the line segment $A_2$, and any connected subset of $A \cup B$ containing $(1,0)$ must contain $b$ or lie entirely within $B_2$. Thus, a connected subset of $A \cup B$ that contains $(1,0)$ and $(1,1)$ must contain both $a$ and $b$; so it is not possible to join both left-hand corners and both right-hand corners at once.

This seems rather counterintuitive; we might have expected that if there was enough material in $A \cup B$ to join pairs of vertices "the hard way", there was enough to join them "the easy way" as well. Indeed, in this case, it is possible to repartition $A \cup B$ into two disjoint connected sets, one containing the upper corners, the other containing the lower corners. Does there exist a "level-crossing" in the unit square—that is, a subset that permits *only* crossed connections of the vertices? More formally:

*Does there exist a subset X of the square which contains disjoint connected sets A and B such that each contains two opposite corners, but that does not contain two disjoint connected sets C and D such that each contains two adjacent corners?*

My personal suspicion is that there does not exist any such subset, but despite sporadic attempts over the last five years, I've been unable to find any reason for this to be so. Can anybody help?

REFERENCES

1. R. C. Buck and E. F. Buck, Advanced Calculus, McGraw-Hill, New York, 1978.
2. F. B. Jones, review of [3], *Math. Rev.*, 34 (1967) #5070.
3. M. Kiang, D. Morrison, and J. Wright, On disjoint connected subsets of a square containing pairs of antipodal points, *Canad. Math. Bull.*, 9 (1966) 631–638.
4. J. A. Seebach and L. A. Steen, Counterexamples in Topology, Springer, New York, 1970.

# NOTES

EDITED BY DENNIS DETURCK, DAVID J. HALLENBECK, AND ANITA E. SOLOW

## On the Square Roots of Infinite Matrices

LILLIAN E. PETERS HUPERT AND ANNE LEGGETT
*Department of Mathematical Sciences, Loyola University, Chicago, Illinois 60626*

**Introduction.** In an address "Matrices I Have Met" at the Joint Mathematics Meetings in New Orleans in January, 1986, Paul Halmos posed the following question: Consider the Cesàro matrix

$$C = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & \cdots \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \cdots \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & \cdots \\ \vdots & & & & & \end{pmatrix}.$$

Is there a matrix $A$ which has the property $A^2 = C$?

We began by constructing lower triangular matrices which have this property. Alerted by a colleague to work of Hausdorff, Toeplitz, and others, we investigated various convergence properties of our square roots. We used computer-generated estimates of the first 25 rows of two of the square roots to formulate conjectures which we then proved on the behavior of their entries. We showed that the first of these square roots takes convergent sequences to convergent sequences and that the second does not even take bounded sequences to bounded sequences. For a while we believed that all the square roots of $C$ were lower triangular; eventually, having found no obstruction, we proved that *all* nonzero infinite matrices have nontriangular square roots. Also, our technique for finding lower triangular square roots of $C$ applies to any infinite lower triangular matrix with all positive entries on the diagonal.

Halmos was, of course, interested in finding a square root of $C$ which is also a bounded operator, i.e., one which takes square-convergent series to square-convergent series. J. B. Conway has proved the existence of a well-defined bounded subnormal operator $\sqrt{C}$ by using the Conway-Olin functional calculus **[1]**, **[2]**. Also, by a theorem of T. Kato **[3]**, $C$ has a bounded accretive square root. The proofs are not constructive, so we do not know a matrix square root which is a bounded subnormal square root.

In our investigation, we used only elementary techniques. The computations are complex, but require knowledge only of calculus, linear algebra, binomial coefficients and, above all, the principle of induction. Most of the proofs are inductive; the square roots are iteratively defined; the inductive relationship between binomial coefficients reflected in Pascal's triangle is essential. In this note we present our results without proof, but will gladly supply details upon request.

We would like to thank our colleagues Martin Buntinas, Cary Huffman, and Gerard McDonald for helpful conversations.

**Constructions of lower-triangular square roots.** One solution to Halmos's question has been known for some time. In 1921, Hausdorff described the following construction in **[4]**:

Let $\nabla$ be the matrix $\nabla = (\delta_{ij})$, where

$$\delta_{ij} = \begin{cases} (-1)^{j-1}\binom{i-1}{j-1} & \text{if } i \geqslant j \\ 0 & \text{if } i < j \end{cases},$$

that is

$$\nabla = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & \\ 1 & -1 & 0 & 0 & 0 & \cdots & \\ 1 & -2 & 1 & 0 & 0 & \cdots & \\ 1 & -3 & 3 & -1 & 0 & \cdots & \\ 1 & -4 & 6 & -4 & 1 & 0 & \cdots \\ & \cdots & & & & & \end{pmatrix},$$

where the nonzero entries are the binomial coefficients with alternating signs. It is straightforward to show that $\nabla = \nabla^{-1}$.

For $\alpha \in R$, let

$$H_\alpha = \nabla \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ 0 & \left(\frac{1}{2}\right)^\alpha & 0 & 0 & \cdots \\ 0 & 0 & \left(\frac{1}{3}\right)^\alpha & 0 & \cdots \\ 0 & 0 & 0 & \left(\frac{1}{4}\right)^\alpha & \cdots \\ \vdots & \vdots & \vdots & & \end{pmatrix} \nabla.$$

Clearly, for $\alpha, \beta \in R$, $H_{\alpha+\beta} = H_\alpha H_\beta$. Because another straightforward computation shows that $H_1$ is the Cesàro matrix, it follows that $H_{1/2}$ has the desired property, i.e., $(H_{1/2})^2 = H_1 = C$.

By a similar construction, we obtain infinitely many lower triangular square roots of $C$. Let $\sigma$ be a function defined on the positive integers such that $\sigma(i) \in \{-1, 1\}$ for all $i$. Let $D_\sigma$ be the diagonal matrix with $n$th diagonal entry $\sigma(n) \cdot 1/\sqrt{n}$. Let $A_\sigma = \nabla D_\sigma \nabla$, the Hausdorff transformation for the diagonal matrix $D_\sigma$. Then it is easy to see that $A_\sigma$ is a lower triangular square root of $C$.

By another straightforward calculation, it can be verified that if $A_\sigma = (a_{ij})$, then

$$a_{ij} = \binom{i-1}{j-1} \sum_{l=0}^{i-j} (-1)^l \sigma(l+j) \frac{1}{\sqrt{l+j}} \binom{i-j}{l} \quad \text{for all } i \geqslant j.$$

This explicit formula is more useful than the iterative definition for proving that various properties hold for $A_\sigma$.

In fact, every lower triangular square root of $C$ is equal to some $A_\sigma$. To see this, we need a more direct method of construction.

Let $B_\sigma$ be a lower triangular matrix defined as follows:

1) for all $i \geqslant 1$, $b_{ii} = \sigma(i)\dfrac{1}{\sqrt{i}}$.

2) for all $i \geqslant 2$, $b_{i,i-1} = 1/i(b_{ii} + b_{i-1,i-1})$.

3) for all $j > 1$ and $i \geqslant j+1$,

$$b_{i,i-j} = \frac{\dfrac{1}{i} - \displaystyle\sum_{k=i-j+1}^{i-1} b_{ik}b_{k,i-j}}{b_{ii} + b_{i-j,i-j}}.$$

The denominators in 2) and 3) are clearly nonzero by 1). Here, we are constructing the matrix $B_\sigma$ by first defining the main diagonal, then the subdiagonal, then the diagonal below that, etc. By straightforward (but tedious) calculation, we show that this iterative method produces all lower triangular square roots of $C$.

THEOREM 1. *Let $B$ be a lower triangular matrix. Then $B^2 = C$ iff $B = B_\sigma$ for some $\sigma\colon Z^+ \to \{1, -1\}$.*

Further, since $B_\sigma$ is completely determined by the entries on its main diagonal, it follows that, for any $\sigma\colon Z^+ \to \{1, -1\}$, $A_\sigma = B_\sigma$.

If $M$ is any infinite lower triangular matrix with positive entries on the diagonal, the method above may easily be modified to produce a square root by choosing $\sigma(i) = 1$ for all $i$. If the diagonal entries are pairwise distinct, then a square root may be defined for each choice of $\sigma$.

**Convergence properties.** Let $A$ be lower triangular. Then $A\vec{x}$ is defined for all $\vec{x}$ by lower triangularity. $A$ is defined to be a **limitation matrix** iff $A\vec{x}$ is bounded whenever $\vec{x}$ is bounded. In [5], [6] it is proved that $A$ is a limitation matrix iff there is a constant $K$ such that $\sum_{j=1}^{\infty}|a_{ij}| < K$ for all $i \in Z^+$. $A$ is defined to be a **regular matrix** iff $A\vec{x}$ and $\vec{x}$ have the same limit whenever $\vec{x}$ is convergent. $A$ is a regular matrix iff the following three conditions hold:

1) $$\sum_{j=1}^{\infty} |a_{ij}| < K \quad \text{for all } i \in Z^+ \text{ for some fixed } K,$$

2) $$\lim_{i\to\infty} a_{ij} = 0 \text{ for all } j \in Z^+, \text{ and}$$

3) $$\lim_{i\to\infty} \sum_{j=1}^{\infty} a_{ij} = 1.$$

(See [5], [6], [7].)

Let $A_+ = A_\sigma$ where $\sigma(i) = 1$ for all $i \in Z^+$, and let $A_{\text{alt}} = A_{\sigma'}$ where $\sigma'(i) = (-1)^{i+1}$ for all $i \in Z^+$. These are the matrices whose entries we approximated on the computer.

The subdiagonal of $A_{\text{alt}}$ blows up in absolute value: it is not difficult to show that $|a_{i,i-1}| \geqslant 2\sqrt{i-1}$. Thus $A_{\text{alt}}$ is not even a limitation matrix.

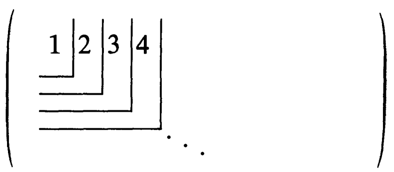
For any $A_\sigma$, by a complicated induction it can be proved that $\sum_{j=1}^{\infty} a_{ij} = a_{11}$ for all $i \in Z^+$. Using the explicit characterization of the entries of $A_+$, induction down the subdiagonals of $A_+$, Pascal's triangle, and the derivatives of $1/\sqrt{x}$, we see that

$a_{ij} \geqslant 0$ for entries $a_{ij}$ of $A_+$. Also, it is easy to see that $a_{ij} < 1/2\sqrt{i}$ for all $i, j \in Z^+$ with $i > j$. Hence we obtain the following theorem:

THEOREM 2. *$A_+$ is a regular matrix.*

Thus the matrix $A_+$, Hausdorff's $H_{1/2}$, is a square root of the Cesàro matrix $C$ which is, like $C$, a regular matrix. $A_{\text{alt}}$, on the other hand, is not even a limitation matrix. Thus the lower triangular square roots of $C$ have strikingly different convergence properties.

**Construction of nontriangular square roots.** In this section we describe a construction of a nontriangular square root $Y$ for any infinite matrix $X$. We build $Y$ in the order indicated by this picture:




The idea is that in section $k$ we correct the $k$th entry of $Y^2$ so that it will agree with the $k$th entry of $X$. We must be careful to ensure that once an entry is corrected, the associated computation is not changed. To this end, we use any counting $\phi$ of $Z^+ \times Z^+$ which has the property that $\phi(r, s) > r, s$ for all $(r, s)$.

Let $X = (x_{ij})$. We construct the matrix $Y = (y_{ij})$ as follows.

Set $y_{11} = 0$. Now, for $k \geqslant 2$, define recursively all entries of $Y$ in the $k$th row and to the left of the diagonal, and in the $k$th column above or on the diagonal.

Let $\phi^{-1}(k) = (r, s)$. For all $i$, $1 \leqslant i \leqslant k$ and $i \neq r$, set $y_{ik} = 0$. For all $j$, $1 \leqslant j \leqslant k$ and $j \neq s$, set $y_{kj} = 0$. If $\sum_{l=1}^{k-1} y_{rl} y_{ls} = x_{rs}$, set $y_{rk} = y_{ks} = 0$. Otherwise, set $y_{rk} = 1$ and $y_{ks} = x_{rs} - \sum_{l=1}^{k-1} y_{rl} y_{ls}$. Notice that, in either case, $\sum_{l=1}^{k} y_{rl} y_{ls} = x_{rs}$.

Further, because $\phi$ is a bijection, we verify that $y_{rl} y_{ls} = 0$ whenever $l > k$. Then $\sum_{l=1}^{\infty} y_{rl} y_{ls} = \sum_{l=1}^{k} y_{rl} y_{ls} = x_{rs}$. Thus we have the following theorem:

THEOREM 3. $Y^2 = X$.

Because the above construction can be used to find a square root of any infinite matrix, it can, in particular, be used to construct a square root of the Cesàro matrix. Moreover, if $X$ is not the zero matrix, it is obvious that the construction always yields a matrix which is not triangular. (In general, the square root will not even be row-finite or column-finite.)

The construction has several additional properties that are worth noting:

1) If the original matrix $X$ has all integer entries, then the square root $Y$ has all integer entries. If $X$ has only rational entries, so does $Y$.

2) The choices of $y_{rk}$ and $y_{ks}$ are clearly not uniquely determined.

3) This construction will work for infinite matrices over any ring $R$ which has the property that, for any $r \in R$, there are $r_1, r_2 \in R$ with $r_1 r_2 = r$. In particular, the construction will work if $R$ has an identity. The ring need not even be associative.

The construction has an interesting extension to finite matrices. Let $R$ be some ring with the property mentioned in 3). If $X$ is an $n \times n$ matrix over $R$, then for

$m = n^2$, there is an $m \times m$ matrix $Y$ over $R$ with square projecting down to $X$, i.e., the $n \times n$ upper left-hand submatrix of $Y^2$ is equal to $X$.

**Open questions.** Is $H_{1/2} = A_+$ the only regular (lower triangular) square root of $C$?

Is Hausdorff's $H_{1/2}$ a bounded subnormal square root of $C$? It is the only obvious candidate.

The nontriangular square root of $C$ constructed has the property that $\Sigma_{j=1}^{\infty} y_{ij} = \infty$. Obviously the $y_{ij}$'s can be defined in such a way that the rows have finite sums. Can the construction be controlled sufficiently to make $Y$ regular?


REFERENCES

1. J. B. Conway and R. F. Olin, A functional calculus for subnormal operators, II, *Memoirs Amer. Math. Soc.*, 184 (1977).
2. J. B. Conway, Subnormal Operators, Pitman Advanced Publishing Program, London, 1981.
3. T. Kato, Perturbation Theory for Linear Operators, Springer-Verlag, New York, 1966.
4. F. Hausdorff, Summationsmethoden und Momentfolgen I, *Math. Zeitschrift*, 9 (1921) 74–109.
5. L. L. Silverman, On the definition of the sum of a divergent series, University of Missouri Studies, Math. Series I (1913) 1–96.
6. O. Toeplitz, Über allgemeine lineare Mittelbildungen, *Prace mat. fiz.*, 22 (1911) 113–120.
7. J. Schur, Über lineare Transformationen in der Theorie der unendlichen Reihen, *J. Reine angew. Math.*, 151 (1920) 79–111.


## A Note on Venn Diagrams


LEWIS PAKULA

*Department of Mathematics, University of Rhode Island, Kingston, RI 02881*


The familiar three circle Venn diagram can be used for graphical solution of logical problems involving three sets $A_1$, $A_2$, $A_3$ because each of the $2^3$ sets $S_1 \cap S_2 \cap S_3$, where $S_i$ is either $A_i$ or $A_i^c$, can be represented as a plane region bounded by arcs of these circles. Venn, who himself referred to this device as "Eulerian circles," observed (without argument, [7]) that 4 circles cannot similarly be used for problems involving 4 sets, though allowing more varied shapes, ellipses say, will work. This observation is perhaps commonplace in introductory discussions of logic (e.g. [**3**, p. 42], [**4**, p. 53]). An elementary demonstration of the inadequacy of circles follows immediately from Problem 44b of [**8**]; Euler's formula for planar graphs can also be used to show this, as well as the fact that ellipses cannot be used for 6 sets [**5**].

A bit of undergraduate linear algebra reveals some general limitations on the use of algebraic curves as Venn diagram boundaries. Let $X$ be a set and $f$ a real function on $X$. Let $p(f) = \{x \in X\colon f(x) > 0\}$. A collection of functions $f_1, \dots, f_k$ will be said to *define boundaries for a k-set Venn diagram* if each of the $2^k$ sets $S_1 \cap \cdots \cap S_k$, with $S_i = p(f_i)$ or $p(-f_i)$, is nonempty.

PROPOSITION 1. *Suppose that $U$ is a vector space of real functions on $X$ with dimension $n$ and $V = \{g + c\colon g \in U,\ c \in R\}$. Then no collection of $n + 1$ functions chosen from $V$ define boundaries for an $(n + 1)$-set Venn diagram.*

*Proof.* Let $f_1, \ldots, f_{n+1} \in V$ and suppose $f_j = g_j + c_j$ with $g_j \in U$. We will show that there is some (strict) orthant of $R^{n+1}$ which does not contain any vector of the form

$$(f_1(x), \ldots, f_{n+1}(x)), \tag{1}$$

with $x \in X$. This is easily seen to imply that the chosen functions cannot define boundaries for a Venn diagram. Because $\dim(U) = n$, there is a nonzero vector $\mathbf{v} = (v_1, \ldots, v_{n+1})$ such that $v_1 g_1 + \cdots + v_{n+1} g_{n+1} = 0$. Clearly $\mathbf{v}$ can be chosen to make $d = v_1 c_1 + \cdots + v_{n+1} c_{n+1} \geqslant 0$. This means that each vector (1) lies in the affine manifold $S = \{\mathbf{y}: \mathbf{v} \cdot \mathbf{y} = -d\}$. Let $O$ be the orthant

$$\{(y_1, \ldots, y_{n+1}): y_i > 0 \quad \text{if} \quad v_i > 0, \quad y_i < 0 \quad \text{if} \quad v_i \leqslant 0\}$$

Then $\mathbf{v} \cdot \mathbf{y} > 0$ for any $\mathbf{y} \in O$. Hence $O \cap S$ is empty and we are done.

**Examples.** Let $X = R^2$, $U = \operatorname{span}\{x^2 + y^2, x, y\}$, and let $V$ be as in Proposition 1. Then any four circles in the plane can be specified by $f_1, \ldots, f_4 \in V$ (via $f_i(x) = 0$) and Proposition 1 implies that these cannot be boundaries for a 4-set Venn diagram.

Venn used 4 ellipses, with principal axes parallel to the $x$ and $y$ axes, as a diagram for 4 sets [**7**, p. 7]. However, by letting $U = \operatorname{span}\{x^2, y^2, x, y\}$, we see that no such ellipses will work for 5 sets. Grünbaum [**5**, p. 13] gives an illustration of how 5 ellipses, with axes unrestricted, can be used for 5 sets, but letting $U = \operatorname{span}\{x^2, y^2, x, y, xy\}$ shows that no such ellipses serve for 6.

This reasoning can be extended to ellipsoids (with variously restricted axes) in $R^n$. For example, the reader can easily construct 4 spheres in $R^3$ which give a "Venn diagram" for 4 sets and then apply Proposition 1 to see that 5 spheres can't similarly be used for 5 sets.

Proposition 1 extends an idea in [**6**]. It is closely related to results in combinatorial geometry which have applications to pattern recognition, approximation theory, and probability theory (see e.g. [**1**], [**2**], [**6**]).

REFERENCES

1. P. Assouad, Densitie et dimension, *Ann. Inst. Fourier* (Grenoble) 33 no. 3 (1983) 233–282.
2. T. M. Cover, Geometrical and statistical properties of systems of linear inequalities with applications to pattern recognition. *IEEE Trans. Electron. Comput.* EC-14 (1965) 326–334.
3. M. Gardner, Logic Machines and Diagrams, McGraw-Hill, 1958.
4. A. W. Goodman and J. S. Ratti, Finite Mathematics with Applications, Macmillan, 1971.
5. B. Grünbaum, Venn diagrams and independent families of sets, *Mathematics Magazine*, 48 (1975) 12–22.
6. H. S. Shapiro, Some negative theorems of approximation theory, *Michigan Math. J.*, 11 (1964) 211–217.
7. J. Venn, On the diagrammatic and mechanical representation of propositions and reasonings, *Philosophical Magazine*, Series 5, vol. 10 (1880).
8. A. M. Yaglom and I. M. Yaglom, Challenging Mathematical Problems with Elementary Solutions, vol. I, Holden-Day, 1964.

# Direct Sum of $J$-Rings and Zero Rings

STEVE LIGH
*Department of Mathematics, University of Southwestern Louisiana, Lafayette, LA 70504*

JIANG LUH
*Department of Mathematics, North Carolina State University, Raleigh, NC 27695*

A ring in which every element $x$ satisfies $x^{n(x)} = x$, for some integer $n(x) > 1$, is called a $J$-ring, and it is well known [1] that such a ring is commutative. In [4], it was shown that a ring $R$ is commutative if it satisfies the following condition:

($*$) For each $x, y \in R$, there exists an integer $n = n(x, y) > 1$ such that $(xy)^n = xy$.

It was pointed out that the above condition is weaker than the condition $x^{n(x)} = x$ for all $x$ in $R$, since there exist non-$J$-rings satisfying ($*$). The example given in [4] was any zero ring $R$, i.e. $xy = 0$ for all $x$ and $y$ in $R$.

The purpose of this note is to show that there are no other examples by establishing the following structure theorem of any ring $R$ satisfying ($*$): $R$ is a direct sum of a $J$-ring and a zero ring. We use the commutativity of $R$ from [4] throughout this note.

LEMMA. *Let $R$ be a ring satisfying condition ($*$). Then $A = \{x \in R\colon x^2 = 0\}$ and $B = \{y \in R\colon y^{n(y)} = y$ for some integer $n(y) > 1\}$ are ideals of $R$.*

*Proof.* If $x \in A$, then $(xy)^2 = x^2y^2 = 0$ implies that $xR = 0$. For any $x, y \in A$, $(x - y)^2 = x^2 - 2xy + y^2 = 0$. Hence $A$ is an ideal of $R$.

For any $x$, $y$ in $B$, there exist integers $n = n(x) > 1$ and $m = m(y) > 1$ such that $x^n = x$ and $y^m = y$. Let $k = (n-1)m - (n-2) = (m-1)n - (m-2)$. Then it is clear that $x^k = x$ and $y^k = y$. Hence, for any $x_1, x_2, \ldots, x_i$ in $B$, there exists $k > 1$ such that $x_t^k = x_t$ for all $t = 1, 2, \ldots, i$.

Next we wish to show for any $x$, $y$ in $B$, $x + y$ is in $B$. Since $(x + y)y$ is in $B$ by ($*$), there exists $k$ such that $x^k = x$, $y^k = y$ and $((x + y)y)^k = (x + y)y$. It follows that

$$\begin{aligned}(x+y)^k y^k &= \left(x^k + \binom{k}{1}x^{k-1}y + \cdots + \binom{k}{k-1}xy^{k-1} + y^k\right)y^k \\ &= (x+y)y \\ &= xy + y^2.\end{aligned}$$

Hence,

$$\binom{k}{1}x^{k-1}y^2 + \cdots + \binom{k}{k-1}xy^k = 0.$$

Now multiplying both sides of the above by $y^{2k-3}$ yields

$$\binom{k}{1}x^{k-1}y + \cdots + \binom{k}{k-1}xy^{k-1} = 0.$$

Hence,

$$(x+y)^k = x^k + \binom{k}{1}x^{k-1}y + \cdots + \binom{k}{k-1}xy^{k-1} + y^k = x + y.$$

Thus for any $x$, $y$ in $B$, $x + y$ is in $B$. It is easy to see that if $x$ is in $B$ and $r$ is in $R$, then $xr$ and $-x$ are in $B$. This shows that $B$ is an ideal of $R$.

Now we are ready to state and prove the result concerning the structure of rings satisfying $(*)$.

THEOREM. *Let $R$ be a ring satisfying condition $(*)$. Then $R = A \oplus B$, where $A$ is a zero ring and $B$ is a $J$-ring.*

*Proof.* If $x \in A \cap B$, then $x^2 = 0$ and $x^n = x$, $n > 1$. Thus $x = 0$.

Let $r \in R$. By $(*)$, $(rr)^n = r^{2n} = r^2$ for some $n > 1$. Now $r = (r - r^{2n-1}) + r^{2n-1}$. Since $(r - r^{2n-1})^2 = r^2 - 2r^{2n} + r^{4n-2} = 0$, it follows that $r - r^{2n-1} \in A$. Now $r^{2n-1}$ is in $B$ since $n > 1$. This completes the proof.

The following corollary was obtained by Lee in [2].

COROLLARY. *A ring $R$ satisfying the condition $(xy)^2 = xy$ for all $x$, $y$ in $R$ is a direct sum of a Boolean ring and a zero ring.*

*Remark* 1. For other descriptions of rings satisfying condition $(*)$, see [3].

*Remark* 2. The above corollary was also obtained by Scott Beslin and Pat Jones.

REFERENCES

1. I. N. Herstein, Non-Commutative Rings, Carus Mathematical Monograph No. 15, Mathematical Association of America, Washington, D.C., 1968.
2. S. M. Lee, Rings and semigroups which satisfy the identity $(xy)^n = xy = x^n y^n$, *Nanta Mathematica*, 6 (1973) 21–28.
3. S. Ligh and Y. Utumi, Direct sum of strongly regular rings and zero rings, *Proc. Japan Acad.*, 50 (1974) 589–592.
4. M. Ó Searcóid and D. MacHale, Two elementary generalizations of Boolean rings, *Amer. Math. Monthly*, 93 (1986) 121–122.

## A Very Short Proof of Stirling's Formula

J. M. PATIN
*Le Bourg, Epineux-Le-Séguin, 53340 Ballee, France*

The purpose of this note is to give an extremely short proof of Stirling's formula which uses only the Lebesgue dominated convergence theorem, as opposed to [1] where I find the use of the Central Limit Theorem or the inversion theorem for characteristic functions unsatisfactory since these results cannot be reasonably judged elementary.

In [2], P. Diaconis and D. Freeman published another proof with a similar idea, but the proof below is clearer and shorter. For $x > 0$

$$\Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}\,dt = 2\int_0^{+\infty} u^{2x-1}e^{-u^2}\,du;$$

hence,

$$\frac{\Gamma(x)e^x\sqrt{x}}{x^x} = 2\int_0^{+\infty} e^{x-u^2}\left(\frac{u}{\sqrt{x}}\right)^{2x-1} du.$$

Changing variables ($u = \sqrt{x} + v$), one gets

$$\frac{\Gamma(x)e^x\sqrt{x}}{x^x} = 2\int_{-\sqrt{x}}^{+\infty} e^{-2vx^{1/2}}\left(1+\frac{v}{\sqrt{x}}\right)^{2x-1} e^{-v^2}\,dv = 2\int_{-\infty}^{+\infty}\varphi_x(v)e^{-v^2}\,dv,$$

where

$$\varphi_x(v) = \begin{cases} 0 & \text{for } v \leqslant -\sqrt{x} \\ e^{-2vx^{1/2}}\left(1+\dfrac{v}{\sqrt{x}}\right)^{2x-1} & \text{for } v \geqslant -\sqrt{x}. \end{cases}$$

First, using the series expansion,

$$\log\left(1+\frac{v}{\sqrt{x}}\right) = \frac{v}{\sqrt{x}} - \frac{v^2}{2x} + \cdots$$

we see that, for a fixed $v$,

$$\log\varphi_x(v) = -v^2 + 0\left(\frac{1}{\sqrt{x}}\right) \text{ as } x \to +\infty; \quad \text{hence } \lim_{x\to+\infty}\varphi_x(v) = e^{-v^2}.$$

Second, $\varphi_x(v)$ has its maximum at

$$v = -\frac{1}{2\sqrt{x}}, \quad \text{hence, } \varphi_x(v) \leqslant e\left(1-\frac{1}{2x}\right)^{2x-1},$$

which remains bounded when $x \to +\infty$, since

$$\lim_{x\to+\infty}\left(1-\frac{1}{2x}\right)^{2x-1} = e^{-1}.$$

Thus one finally has

$$\begin{cases} \lim\limits_{x\to+\infty}\varphi_x(v)e^{-v^2} = e^{-2v^2} \\ \varphi_x(v)e^{-v^2} = 0(e^{-v^2}) \text{ as } x \to +\infty. \end{cases}$$

and, bearing in mind Walter Rudin's remark [**3**, p. 194, 9.3(a)], which explains how the Lebesgue dominated convergence theorem [**3**, p. 27], can be applied to an uncountable family such as $\varphi_x(v)e^{-v^2}$, we conclude:

$$\lim_{x\to+\infty}\frac{\Gamma(x)e^x\sqrt{x}}{x^x} = 2\int_{-\infty}^{+\infty} e^{-2v^2}\,dv = 2\sqrt{\frac{\pi}{2}} = \sqrt{2\pi}.$$

REFERENCES

1. C. R. Blyth and P. K. Pathak, A note on easy proofs of Stirling's theorem, this MONTHLY, 93 (1986) 376–379.
2. P. Diaconis and D. Freeman, An elementary proof of Stirling's formula, this MONTHLY, 93 (1986) 123–125.
3. W. Rudin, Real and Complex Analysis, 2nd ed., McGraw-Hill, New York, 1974.

EDITED BY MELVIN HENRIKSEN AND STAN WAGON

# A Euclidean Model for Euclidean Geometry

ADOLF MADER
*Department of Mathematics, University of Hawaii, Honolulu, HI 96822*

**I. Introduction.** In courses on non-Euclidean geometry models of hyperbolic geometry play an important role. The standard models are Klein's disk model and the disk model and half-plane model named after Poincaré [**1**], [**4**], [**6**], [**7**], [**8**]. In each case the model is embedded in the Euclidean plane and the concepts of hyperbolic geometry are interpreted in Euclidean terms. One wonders whether there is a similar model of Euclidean geometry within Euclidean geometry, and in particular a model that is bounded as a Euclidean object. Then for once one could see the whole Euclidean world that tends to run off the blackboard all too fast. A little reflection yields the correct answer: Of course! After all, all that is needed is a bijection of the standard Euclidean space with a bounded part of itself, and by means of this bijection the structure of Euclidean space can be transferred to the model. Doing this randomly will likely result in a model which is quite intractable without going back to the original. We will below propose a model that is reasonably simple in itself and has a transparent relationship with standard Euclidean space. It demonstrates the difference between the logical content of an axiom system and its interpretations; in particular it destroys the faith in a preordained single model of Euclidean space. It can serve as an introduction to the idea of models of a geometry, and thus make the models of unfamiliar geometries more palatable. It further constitutes a source of questions and exercises, and looks like a pedagogically convenient way to introduce the projective extension of Euclidean space.

While there was never a doubt that this model must have appeared someplace before, I only discovered it in the textbook by David Gans [**3**] when I revised my article. Gans uses the model extensively for motivating and illustrating projective geometry. It is introduced on page 212 after the study of transformations of various kinds. However, the model can be used profitably from the start in any geometry course and deserves more publicity. I would like to thank the referee for a thorough review and many helpful suggestions.

**II. The Model $\mathbb{F}$.** Let $\mathbb{F}$ be the interior of a circle $\omega$ of radius $r$ in the standard Euclidean plane $\mathbb{E}$. The straight lines in $\mathbb{F}$ are the half-ellipses of $\mathbb{E}$ whose major axes are diameters of $\omega$ and the diameters themselves. It is convenient to include among the "half-ellipses" the diameters of $\omega$, which then coincide with their "major axes" and have degenerate (one-point) "minor axes." At this stage we may observe:

(2.1) The families of parallel lines in $\mathbb{F}$ are the families of half-ellipses with common major axes.

The angle measure between two straight lines of $\mathbb{F}$ is the $\mathbb{E}$-measure of the angle between their major axes. The $\mathbb{F}$-distance $\bar{d}$ between the center $O$ of $\omega$ and a point

of $\mathbb{F}$ with $\mathbb{E}$-distance $d$ from $O$ is

$$\bar{d} = rd/\sqrt{r^2 - d^2}. \tag{2.2}$$

The distance between any two points on a diameter is computed from their distances from $O$, and the distance between two arbitrary points can be computed by transferring it to a diameter by means of a parallelogram.

2.3 THEOREM. *$\mathbb{F}$ is a model of the Euclidean plane.*

The validity of 2.3 is established by means of the following map.

(2.4) The bijection $\sigma$: $\mathbb{E} \to \mathbb{F}$. The plane $\mathbb{E}$ is contained in 3-dimensional Euclidean space and we choose Cartesian coordinates in the latter such that $\mathbb{E}$ is the $x$–$y$ plane, $\omega$ is the circle $x^2 + y^2 = r^2$, and $\mathbb{F}$ is the interior of $\omega$. Let $S$ be the hemisphere $x^2 + y^2 + (z - r)^2 = r^2$, $z < r$, with center $C$. For the initiated, $\sigma$ is simply the central or gnomonic projection from $\mathbb{E}$ onto $S$ composed with the orthogonal projection of $S$ onto $\omega$ [1], [2]. Explicitly, given $P \in \mathbb{E}$, let $P_1$ be the point of intersection of the line $PC$ with $S$ and let $\bar{P} = \sigma(P)$ be the intersection with $\mathbb{E}$ of the parallel to the $z$-axis through $P_1$.

It is clear that this map $\sigma$: $\mathbb{E} \to \mathbb{F}$ is a bijection. The model $\mathbb{F}$ is established by using $\sigma$ to transfer to it the structure of $\mathbb{E}$. The image $\sigma(a)$ of a straight line $a$ of $\mathbb{E}$ is obtained by intersecting the plane through $a$ and $C$ with $S$ which results in a great semicircle $a_1$. Projecting $a_1$ into $\mathbb{F}$ yields an ellipse whose principal axis is the projection of the diameter of $a_1$ in $z = r$ (Fig. 1). It is now clear that angles in $\mathbb{F}$ are measured as described above. In terms of coordinates $\sigma$ is given as follows.

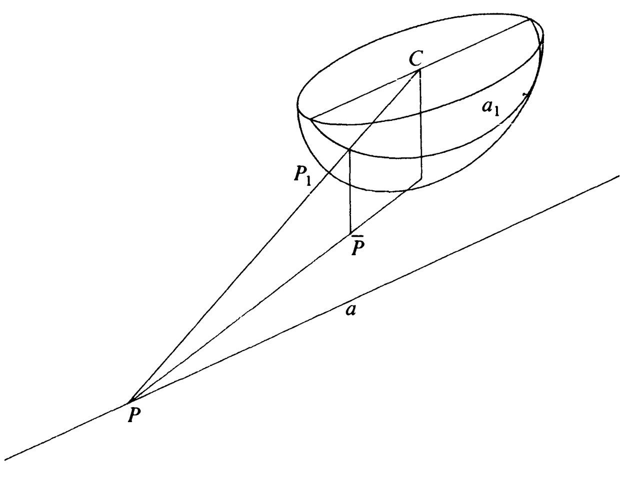



FIG. 1

(2.5) Let $P$ be a point of $\mathbb{E}$ and let $\bar{P} = \sigma(P)$. If $(\rho, \varphi)$ are the polar coordinates of $P$ and $(\bar{\rho}, \bar{\varphi})$ the polar coordinates of $\bar{P}$ then

$$\bar{\rho} = r\rho\Big/\sqrt{r^2 + \rho^2}, \qquad \bar{\varphi} = \varphi,$$
$$\rho = r\bar{\rho}\Big/\sqrt{r^2 - \bar{\rho}^2}, \qquad \varphi = \bar{\varphi}.$$

If $(x, y)$ are the Cartesian coordinates of $P$ and $(\bar{x}, \bar{y})$ the Cartesian coordinates of $\bar{P}$ then

$$\bar{x} = rx\Big/\sqrt{r^2 + x^2 + y^2}, \qquad \bar{y} = ry\Big/\sqrt{r^2 + x^2 + y^2},$$
$$x = r\bar{x}\Big/\sqrt{r^2 - \bar{x}^2 - \bar{y}^2}, \qquad y = r\bar{y}\Big/\sqrt{r^2 - \bar{x}^2 - \bar{y}^2}.$$

*Proof.* In Fig. 2, $OP = \rho$, $O\bar{P} = \bar{\rho}$ and $\bar{P}P_1 = r - \sqrt{r^2 - \bar{\rho}^2}$ from the equation of the sphere. Similar triangles yield

$$\frac{\rho}{r} = \frac{\rho - \bar{\rho}}{r - \sqrt{r^2 - \bar{\rho}^2}},$$

whence $\rho = r\bar{\rho}\Big/\sqrt{r^2 - \bar{\rho}^2}$. The rest follows easily.

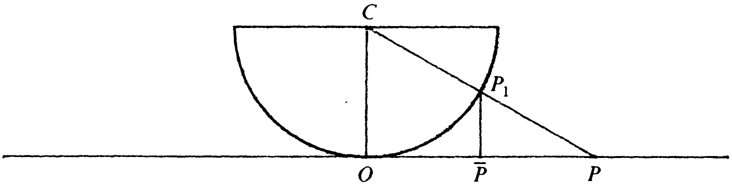



FIG. 2

The formulae (2.5) together with the fact that $\mathbb{F}$ is a Euclidean geometry (being a carbon copy of $\mathbb{E}$) show that the distance measures in $\mathbb{F}$ are obtained as claimed, and 2.3 is now proven.

Another model of Euclidean geometry has appeared partway through the above proof. The central projection maps the Euclidean plane $\mathbb{E}$ bijectively onto the hemisphere

$$\mathbb{H}: x^2 + y^2 + (z - r)^2 = r^2, \qquad z < r,$$

transforming straight lines into great semicircles. The model $\mathbb{F}$ may be viewed as a perspective picture of the hemispherical model $\mathbb{H}$.

Translating properties of $\mathbb{F}$ into the language of $\mathbb{E}$ yields facts and construction problems such as the following.

(2.6) Given a circle $\omega$ and two points in its interior, there exists a unique ellipse passing through the given points and having as major axis a diameter of $\omega$. Find a straightedge and compass construction of the major and minor axis of this ellipse.

(2.7) Given a circle $\omega$, a diameter of $\omega$, and a point in the interior of $\omega$, there exists a unique ellipse having the given diameter as major axis and passing through the given point. Find a straightedge and compass construction of the minor axis of this ellipse.

(2.8) Let $A$ be an ellipse whose major and minor axes are given. Let $a$ be a line passing through the center of $A$. Find a straightedge and compass construction of the points of intersection of $a$ and $A$.

(2.9) Let $A_1$ and $A_2$ be concentric ellipses with known major and minor axes. If the major axes of $A_1$ and $A_2$ have equal lengths, find a straightedge and compass construction of the points of intersection of $A_1$ and $A_2$.

These construction problems can be solved by mapping the $\mathbb{F}$-lines back to $\mathbb{E}$ and thereby turning the ellipses into straight lines. Figure 3 indicates how $P$ can be constructed from $\bar{P} = \sigma(P)$ and conversely.

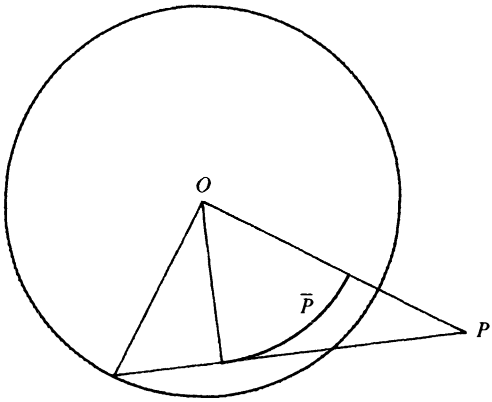



FIG. 3

**III. The Global Picture.** It is now possible to illustrate Euclidean curves in their entirety in $\mathbb{F}$. FIGURES 4–7 illustrate the kinds of graphs one gets in $\mathbb{F}$. In each figure the radius of $\mathbb{F}$ is 12 units.

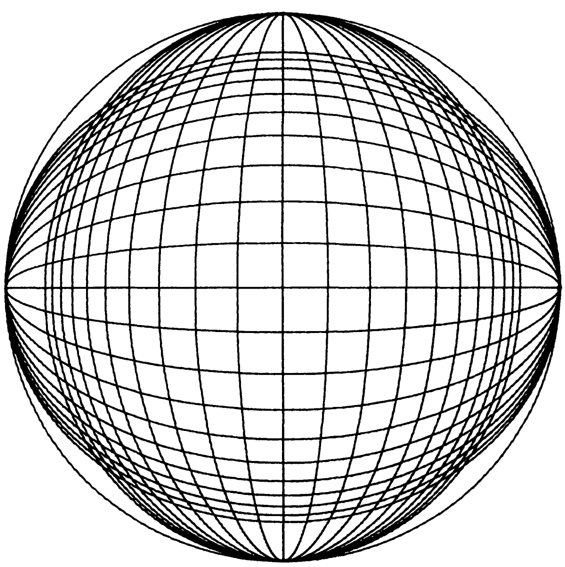


FIG. 4. Coordinate grid $x = 2n$, $y = 2n$, $-10 \leqslant n \leqslant +10$.

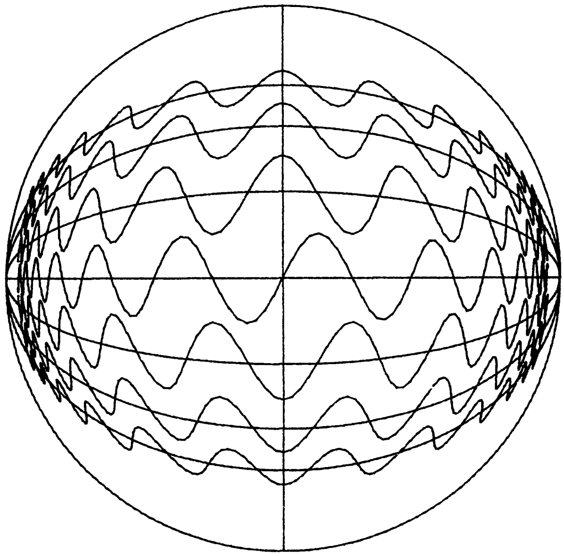


FIG. 5. Sine and cosines $y = 2\sin x$, $y = \pm 2(2n + \cos x)$ $y = \pm 4n, 1 \leqslant n \leqslant 3$

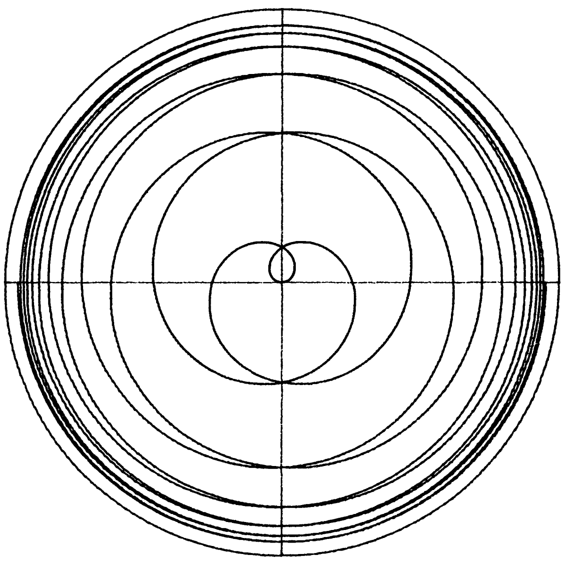


FIG. 6. The Archimedean spiral $\rho = \varphi$.

As in the case of the models of hyperbolic geometry the model $\mathbb{F}$ is deceptive in suggesting a distinguished center of the Euclidean plane. This center is distinguished only from the point of view of the ambient space but indistinguishable from other points from within the model.

The coordinate grid (FIG. 4) is readily recognized as the perspective picture of the corresponding family of great semicircles of the hemispherical model $\mathbb{H}$. It is a bit more challenging to imagine the preimages on the hemisphere of the other figures.

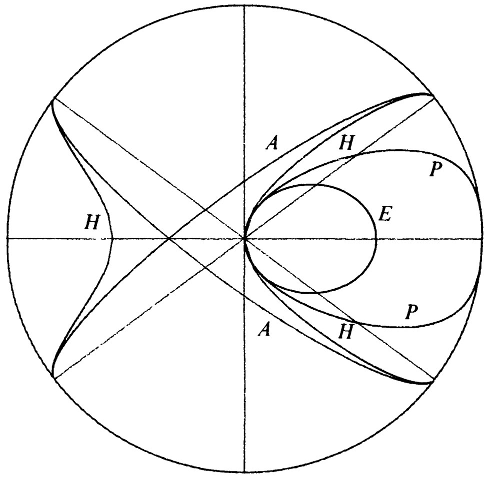



FIG. 7. Conic sections.

E: Ellipse $\dfrac{(x-4)^2}{16} + \dfrac{y^2}{9} = 1$, P: {Parabola $y^2 = 4x$

H: Hyperbola $\dfrac{(x+4)^2}{16} - \dfrac{y^2}{9} = 1$ with asymptotes A: $y = \pm \dfrac{3}{4}(x+4)$

**IV. Projective Geometry.** The ideal points that must be added to $\mathbb{F}$ in order to get the real projective plane $\mathbb{P}$ are there for all to see. They are the points of $\omega$, and $\omega$ is the line at infinity. A family of parallels meets at the endpoints of their common principal axes, and at least at this stage we see that the two endpoints must be identified in order to preserve the incidence axioms. Having these ideal points available, Fig. 7 suggests that the distinguishing features of ellipses, parabolas, and hyperbolas are merely that they have none, one, or two points in common with the ideal line $\omega$. Asymptotic curves meet on $\omega$ but the converse need not be true as we can see from the parabola.

The projective closure of the hemispherical model $\mathbb{H}$ is the closed hemisphere

$$x^2 + y^2 + (z-r)^2 = r^2, \qquad z \leqslant r,$$

with antipodal points on the equator identified. In this setting an angle and distance measure can be defined conveniently [1], [6], [7] yielding a standard model of elliptic geometry. More models of various geometries can be found using the full sphere and utilizing stereographic projection [1], [5].

**V. Extension to Higher Dimensions.** There is no difficulty at all to extending the isomorphism (2.5) to arbitrary dimension $n$. If $\alpha_1, \alpha_2, \dots, \alpha_n$ denote direction angles, $\rho, \bar{\rho}$ the distances from the origin, and $x_i, \bar{x}_i$ the Cartesian coordinates of corresponding points $P$ and $\bar{P}$ respectively then the transformation equations are:

$$\bar{\rho} = \frac{r\rho}{\sqrt{r^2-\rho^2}}, \qquad \bar{\alpha}_i = \alpha_i; \qquad \rho = \frac{r\bar{\rho}}{\sqrt{r^2-\bar{\rho}^2}}, \qquad \alpha_i = \bar{\alpha}_i. \tag{5.1}$$

$$\bar{x}_i = \frac{rx_i}{\sqrt{r^2+\Sigma x_i^2}}; \qquad x_i = \frac{r\bar{x}_i}{\sqrt{r^2-\Sigma \bar{x}_i^2}}. \tag{5.2}$$

The representative hyperplane $x_1 = A$ is transformed into the half-ellipsoid $(r^2 + A^2)x_1^2 + A^2x_2^2 + \cdots + A^2x_n^2 = A^2r^2$, $x_1 \cdot A \geqslant 0$.

**VI. Philosophical Implications.** We all know the murals in churches in which the deity is surrounded by angels and looking down onto earth from his seat on the clouds. It seems like a rather naive image now when every child knows that the immensity of space only begins beyond the clouds, and we expect to encounter strange creatures of somewhat human forms on space ventures rather than God and angels. However, slight corrections in the traditional pictures could create depictions which are mathematically perfectly consistent: simply enter the impenetrable shell of ideal points between the world and the heavens. Even in a Euclidean universe there is plenty of space for one or many worlds like ours and still more for heaven and hell—it all depends on how you measure.

REFERENCES

1. H. S. M. Coxeter, Introduction to Geometry, John Wiley & Sons, New York, 5th Printing, 1966.
2. H. S. M. Coxeter and S. L. Greitzer, Geometry Revisited, New Mathematical Library No. 19, Mathematical Association of America, Washington, 1967.
3. D. Gans, Transformations and Geometries, Appleton-Century-Crofts, New York, 1969.
4. M. J. Greenberg, Euclidean and Non-Euclidean Geometries, 2nd ed., W. H. Freeman and Company, San Francisco, 1980.
5. N. W. Johnson, Absolute polarities and central inversions, in The Geometric Vein: The Coxeter Festschrift, C. Davis, B. Grünbaum, F. A. Sherk, eds., Springer Verlag, New York, 1981, 443–464.
6. H. Meschkowski, Non-Euclidean Geometry, Academic Press, New York and London, 1964.
7. E. Moise, Elementary Geometry from an Advanced Standpoint, Addison-Wesley, Reading, Mass. 1963.
8. C. R. Wylie, Jr., Foundations of Geometry, McGraw-Hill, New York, 1964.

# Orthogonal Bases of $\mathbb{R}^3$ with Integer Coordinates and Integer Lengths

ANTHONY OSBORNE AND HANS LIEBECK
*Department of Mathematics, University of Keele, Staffordshire ST5 5BG, England*

An interesting problem that arises when studying orthogonal bases of $\mathbb{R}^3$ is to find such bases of vectors with integer coordinates and integer lengths. The basis $\{(2,2,-1), (2,-1,2), (-1,2,2)\}$ and its relatives feature prominently in textbooks, but there is a shortage of other examples. In this note we give a complete solution of the problem.

It clearly suffices to consider the *primitive* bases $\{\mathbf{u},\mathbf{v},\mathbf{w}\}$ of $\mathbb{R}^3$ where each vector $\mathbf{u},\mathbf{v},\mathbf{w}$ has integer length and the three coordinates of $\mathbf{u},\mathbf{v},\mathbf{w}$ are, in each case, relatively prime integers. The general solution of our problem is then obtained from the bases $\{a\mathbf{u}, b\mathbf{v}, c\mathbf{w}\}$, where $a$, $b$ and $c$ are arbitrary nonzero integers.

**A simple construction.** The following simple construction provides a rich source of orthogonal bases of the type we want. Consider nonzero vectors in $\mathbb{R}^3$ with

integer elements $p$, $q$ and $r$, not all zero, and let

$$S = \{\mathbf{u} = (p, q, r), \mathbf{v} = (q, r, p), \mathbf{w} = (r, p, q)\}.$$

The condition for $\mathbf{u}$ and $\mathbf{v}$ to be orthogonal is that

$$pq + qr + rp = 0, \tag{1}$$

and this condition is necessary and sufficient for the vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ to be pairwise orthogonal. Moreover, if (1) holds then

$$p^2 + q^2 + r^2 = (p + q + r)^2, \tag{2}$$

so the vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ all have integer length $|p + q + r|$. We have proved the following result.

THEOREM 1. *Let $p$, $q$ and $r$ be integers, not all zero. Then $\{\mathbf{u} = (p, q, r)$, $\mathbf{v} = (q, r, p)$, $\mathbf{w} = (r, p, q)\}$ is an orthogonal basis of $\mathbb{R}^3$ if and only if condition* (1) *holds. Moreover, if this condition is satisfied then the vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ all have length $|p + q + r|$.*

In view of Theorem 1, it is interesting to obtain the general solution of (1) in integers. Notice that if $p + q = 0$ then (1) is completely solved by $p = q = 0$, $r$ arbitrary. This case is excluded in the following theorem.

THEOREM 2. *The general solution in integers of $pq + qr + rp = 0$ with $p + q \neq 0$ is given by*

$$p = a(a + b)k, \qquad q = b(a + b)k, \qquad r = -abk$$

*for integers $a$, $b$ and $k$ with* $\gcd(a, b) = 1$. *Moreover,* $\gcd(p, q, r) = 1$ *if and only if* $k = \pm 1$.

*Proof.* Let $p, q, r$ be integers satisfying (1) with $p + q \neq 0$. Then

$$r = \frac{-pq}{p + q}, \tag{3}$$

and this is an integer. Let $d = \gcd(p + q, p)$. Then also $d = \gcd(p + q, q)$. Put

$$p = da, \qquad q = db. \tag{4}$$

Then $\gcd(a + b, a) = \gcd(a + b, b) = 1$, hence also $\gcd(a, b) = 1$. Condition (3) implies that $(a + b)r = -dab$. Since $a + b$ is relatively prime to both $a$ and $b$, this means that $a + b$ divides $d$. So $d = (a + b)k$ for some integer $k$. Substitution in (4) leads to the required result.

As a consequence of Theorem 1 we obtain a further supply of orthogonal bases of $\mathbb{R}^3$ with integer coordinates and integer lengths as follows.

Let $\mathbf{u} = (p, q, r)$ be a nonzero vector with integer coordinates $p$, $q$ and $r$ satisfying condition (1) and with $\gcd(p, q, r) = 1$. The problem is to construct suitable vectors $\mathbf{s}$ and $\mathbf{t}$ such that $\{\mathbf{u}, \mathbf{s}, \mathbf{t}\}$ is an orthogonal basis of $\mathbb{R}^3$ of the desired type. By Theorem 1 the orthogonal complement of span $\{\mathbf{u}\}$ in $\mathbb{R}^3$ is spanned by the vectors $\mathbf{v} = (q, r, p)$ and $\mathbf{w} = (r, p, q)$, so $\mathbf{s}$ is orthogonal to $\mathbf{u}$ if and only if there exist $x, y \in \mathbb{R}$ such that

$$\mathbf{s} = x(q, r, p) + y(r, p, q).$$

If $\mathbf{s}$ has integer coordinates then $x$ and $y$ must be rational, and so some integer multiple of $\mathbf{s}$ is defined as a sum of integer multiples of $(q, r, p)$ and $(r, p, q)$. Therefore there is no loss of generality if we assume that $x$ and $y$ are integers. However, the basis of $\mathbb{R}^3$ that we obtain containing $\mathbf{u}$ and $\mathbf{s}$ may turn out to be nonprimitive.

An easy calculation shows that

$$\|\mathbf{s}\|^2 = l^2(x^2 + y^2)$$

where $l = |p + q + r|$, the length of $\mathbf{u}$. Thus $\mathbf{s}$ has integer length if and only if $(|x|, |y|, \sqrt{x^2 + y^2})$ is a Pythagorean triple of integers. (We admit the possibility that $x$ or $y$ is zero.) The vector

$$\mathbf{t} = x(r, p, q) - y(q, r, p)$$

completes an orthogonal basis $\{\mathbf{u}, \mathbf{s}, \mathbf{t}\}$ of $\mathbb{R}^3$, for clearly $\mathbf{u} \cdot \mathbf{t} = \mathbf{s} \cdot \mathbf{t} = 0$. Also

$$\|\mathbf{t}\|^2 = l^2(x^2 + y^2) = \|\mathbf{s}\|^2,$$

so $\{\mathbf{u}, \mathbf{s}, \mathbf{t}\}$ is a basis of the desired type if $(|x|, |y|, \sqrt{x^2 + y^2})$ is a Pythagorean triple of integers. Apart from the possible cancellation of a common factor, all integer length orthogonal bases with integer entries that contain $\mathbf{u}$ as a member are obtained in this way.

**Examples.** (a) Consider $\mathbf{u} = (2, 2, -1)$, $x = 3$ and $y = -4$. We obtain the primitive orthogonal basis $\{\mathbf{u}, \mathbf{s} = (10, -11, -2),\ \mathbf{t} = (5, 2, 14)\}$ with vectors of integer lengths 3, 15 and 15, respectively. (b) The construction always leads to an orthogonal basis containing at least two vectors of the same length. However, as we have pointed out above, such a basis need not be primitive. For example, the case $\mathbf{u} = (12, 4, -3)$, $x = 3$ and $y = 4$ leads to the orthogonal basis $\{\mathbf{u}, \mathbf{s} = (0, 39, 52),\ \mathbf{t} = (-25, 48, -36)\}$ of integer lengths 13, 65 and 65. From this we obtain the primitive basis $\{(12, 4, -3), (0, 3, 4), (-25, 48, -36)\}$ of lengths 13, 5 and 65.

The above analysis affords the means of constructing all orthogonal bases of $\mathbb{R}^3$ with integer coordinates and integer lengths that contain a vector $(p, q, r)$ satisfying $pq + qr + rp = 0$. Now let

$$A = \begin{bmatrix} p & q & r \\ s_1 & s_2 & s_3 \\ t_1 & t_2 & t_3 \end{bmatrix}$$

be a matrix whose rows represent such a primitive basis. Then further primitive orthogonal bases of $\mathbb{R}^3$ with integer coordinates and integer lengths are obtained by interchanging two rows or two columns, or by multiplying any row(s) or any column(s) by $-1$. In this way we obtain from Example (a) above the primitive orthogonal basis $\{(2, 2, 1), (10, -11, 2), (-5, -2, 14)\}$. None of these vectors satisfies the condition $pq + qr + rp = 0$. Nevertheless here again we are dealing with a very special case.

**The general construction.** We now start with an arbitrary nonzero vector $\mathbf{u} = (p, q, r)$ with integer coordinates and integer length $l$, and we assume no further conditions on the integers $p$, $q$ and $r$ except that $\gcd(p, q, r) = 1$. The following important result provides a formula for $p$, $q$, $r$ and $l$.

THEOREM 3 [**1**, p. 12]. *The general solution in integers of the equation*

$$p^2 + q^2 + r^2 = l^2 \tag{5}$$

*where* $\gcd(p, q, r) = 1$ *and* $l > 0$ *is given by*

$$sp = 2ac, \qquad sq = 2bc, \qquad sr = -a^2 - b^2 + c^2, \qquad sl = a^2 + b^2 + c^2, \tag{6}$$

*where* $a, b, c$ *are arbitrary integers with* $\gcd(a, b, c) = 1$ *and* $s > 0$ *is taken so that* $\gcd(p, q, r) = 1$.

*Proof.* Suppose first that $p, q, r, l$ is a solution in integers of equation (5) with $\gcd(p, q, r) = 1$. The case $p = q = 0$, $r = \pm 1$, $l = 1$ is trivial, so assume that $p$ and $q$ are not both zero. Let $d = \gcd(p, q, l + r)$. Then $\gcd(d, l - r) = 1$ if $p$ or $q$ is odd, and $\gcd(d, l - r) = 2$ if $p$ and $q$ are both even, for otherwise some prime number would divide $p$, $q$ and $r$. Put

$$p = ad, \qquad q = bd, \qquad l + r = cd \tag{7}$$

so $\gcd(a, b, c) = 1$. Then by (7) and (5)

$$(-a^2 - b^2 + c^2)d^2 = 2r^2 + 2lr = 2rcd \tag{8}$$

and

$$(a^2 + b^2 + c^2)d^2 = 2l^2 + 2lr = 2lcd. \tag{9}$$

Hence,

$$(a^2 + b^2)d^2 = cd(l - r),$$

and so $d$ divides $c(l - r)$. Since $\gcd(d, l - r) = 1$ or 2, it follows that $d$ divides $2c$.
Put $s = 2c/d$. Then $s$ is an integer, and since $l + r$ is positive, $s > 0$. The relations (6) now follow from (7), (8) and (9).
The converse is easily verified, and we omit the proof.

**Example.** Take $(a, b, c) = (4, -7, 13)$. Then in formula (6) $s = 26$, $(p, q, r) = (4, -7, 4)$ of length $l = 9$.

**Remark.** Putting $b = 0$ in Theorem 3 gives a general formula for Pythagorean triples of integers $p, r, l$, where $\gcd(p, r) = 1$.

Now let $\mathbf{u} = (p, q, r)$ be an arbitrary nonzero vector with integer coordinates and integer length. We wish to find another such vector $\mathbf{s}$ that is orthogonal to $\mathbf{u}$. By Theorem 3, $\mathbf{s}$ is a rational multiple of $(2ac, 2bc, -a^2 - b^2 + c^2)$ where $a, b, c$ are integers with $\gcd(a, b, c) = 1$ which are so chosen that $\mathbf{s}$ is orthogonal to $\mathbf{u}$. Setting aside the trivial case when $c = 0$ solves the problem, assume $c \neq 0$, and put $\lambda = a/c$, $\mu = b/c$. Then $\mathbf{s}$ is a rational multiple of the vector with rational coefficients

$$\mathbf{s}^* = (2\lambda, 2\mu, 1 - \lambda^2 - \mu^2).$$

The orthogonality condition $\mathbf{u} \cdot \mathbf{s}^* = 0$ gives

$$2p\lambda + 2q\mu + r(1 - \lambda^2 - \mu^2) = 0,$$

which when viewed as a quadratic equation in $\lambda$ becomes

$$r\lambda^2 - 2p\lambda + (r\mu^2 - 2q\mu - r) = 0. \tag{10}$$

If $r = 0$, equation (10) becomes $-2p\lambda - 2q\mu = 0$. Thus a primitive solution is $\mathbf{s}^* = (q, -p, 0)$. If $r \neq 0$, solutions with $\lambda$ rational require that the discriminant

$$p^2 - r(r\mu^2 - 2q\mu - r) \tag{11}$$

is the square of a rational number $\nu$. From (11) we obtain

$$(q - r\mu)^2 + \nu^2 = l^2 \tag{12}$$

where $l = \|\mathbf{u}\|$, and $\mu$ and $\nu$ are rational numbers.

Let $(|x|, |y|, z)$ with $z > 0$ be a Pythagorean triple. (We again admit the possibility that $x$ or $y$ is zero.) Then (12) is solved in rationals by

$$(q - r\mu) = \frac{xl}{z}, \qquad \nu = \frac{yl}{z} \tag{13}$$

and all solutions of (12) in rationals are of this form. From (13) and (10) we find rationals

$$\lambda = \frac{zp + yl}{zr}, \qquad \mu = \frac{zq - xl}{zr} \tag{14}$$

(without loss of generality) such that the vector $\mathbf{s}^*$ is orthogonal to $\mathbf{u}$. Then also $\mathbf{s}$ is orthogonal to $\mathbf{u}$ and $\{\mathbf{u}, \mathbf{s}, \mathbf{u} \times \mathbf{s}\}$ is an orthogonal basis of $\mathbb{R}^3$ with integer coordinates and integer lengths.

**Examples.** (a) Suppose $\mathbf{u} = (4,7,4)$ of length $l = 9$. Let $x = 3$, $y = -4$, $z = 5$. Then $5\mathbf{s}^* = (-8,4,1)$. The vectors $\mathbf{u} = (4,7,4)$, $\mathbf{s} = (-8,4,1)$ and $\frac{1}{9}(\mathbf{u} \times \mathbf{s}) = (-1,-4,8)$ form a primitive orthogonal basis of $\mathbb{R}^3$. The basis vectors all have length 9. The choice of $x = 3$, $y = 4$, $z = 5$ in (14) leads to the primitive orthogonal basis $\{(4,7,4), (28,4,-35), (-29,28,-20)\}$ with vectors of lengths 9, 45, 45, respectively.

(b) Now suppose $\mathbf{u} = (12,12,1)$ of length $l = 17$. Taking $x = -5$, $y = -12$, $z = 13$ we obtain the primitive orthogonal basis $\{(12,12,1), (-48,241,-2316), (-97,96,12)\}$ with vectors of distinct lengths 17, 2329 and 137. Such primitive orthogonal bases of vectors with distinct lengths appear to be somewhat elusive!

The above technique produces all primitive orthogonal bases containing a given vector $\mathbf{u}$.

**Added in proof.** Since this article went to press, the authors have obtained a method of generating all orthogonal bases of $\mathbb{R}^n$ with integer coordinates and integer lengths by using a matrix approach [**2**].

REFERENCES

1. L. J. Mordell, Diophantine Equations, Academic Press, London, 1969.
2. Hans Liebeck and Anthony Osborne, The generation of all rational orthogonal matrices, submitted to this MONTHLY.

# PROBLEMS AND SOLUTIONS



*For instructions about submitting **proposed** problems for publication in this department see the inside front cover. Please include solutions, relevant references, etc.*

*An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.*

***Solutions** should be sent to the address given on the inside front cover. Two copies suffice.*

*A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a* MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the colution if you can.*

*For instructions about submitting solutions of Problems, which should be mailed before May 30, 1989, see the inside front cover. Please place the solver's name and mailing address on each (doubled-spaced) sheet. Include a self-addressed card or label if an acknowledgement is desired.*

## ELEMENTARY PROBLEMS

E 3301. *Proposed by Hugh L. Montgomery, University of Michigan, Ann Arbor, and Jeffrey D. Vaaler, University of Texas, Austin.*

Suppose $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N$ are vectors in $R^M$, where $N \geq M$. Prove that the set of vectors in $R^M$ expressible in the form

$$\lambda_1\mathbf{a}_1 + \lambda_2\mathbf{a}_2 + \cdots + \lambda_N\mathbf{a}_N, \qquad 0 \leqslant \lambda_i \leqslant 1,$$

has $M$-dimensional volume

$$\sum_I |\det A_I|,$$

where $I$ runs over the subsets of $\{1, 2, \ldots, N\}$ of cardinality $M$ and $A_I$ denotes the $M$ by $M$ matrix whose columns are the $\mathbf{a}_i$ for which $i \in I$.

E 3302. *Proposed by Jim Delany, California Polytechnic State University, San Luis Obispo, CA.*

The mean and standard deviation of any 7 consecutive integers are both integers. What natural numbers greater than 1 share this property with 7?

E 3303. *Proposed by Donald E. Knuth, Stanford University, CA.*

Suppose $m$ and $n$ are positive integers. Find a closed-form expression for

$$\sum_{k=0}^{2^{mn}-1} \binom{k^m}{n} (-1)^{\nu(k)},$$

where $\nu(k)$ is the number of 1's in the binary representation for $k$.

E 3304. *Proposed by Dante Vialetto, Olgiate Olona, Italy.*

Prove that there are infinitely many solutions of

$$ax^2 + by^2 = cz^2$$

in integers $a, b, c, x, y, z$ with

$$a < b < c < x < y < z,$$

and obtain a family of such solutions with as many independent parameters as possible.

E 3305. *Proposed by M. S. Klamkin, University of Alberta, Edmonton.*

If $a, b, c$ are the sides of a triangle with given semiperimeter $s$, determine the maximum values of
(i) $(b-c)^2 + (c-a)^2 + (a-b)^2$,
(ii) $|(b-c)(c-a)| + |(c-a)(a-b)| + |(a-b)(b-c)|$, and
(iii) $(b-c)^2(c-a)^2(a-b)^2$.

E 3306. *Proposed by J. B. Wilker, Scarborough College, University of Toronto, Canada.*

(a) Prove that if $0 < x < \frac{\pi}{2}$, then

$$\left(\frac{\sin x}{x}\right)^2 + \frac{\tan x}{x} > 2.$$

(b) Find the largest constant $c$ such that

$$\left(\frac{\sin x}{x}\right)^2 + \frac{\tan x}{x} > 2 + cx^3 \tan x$$

for $0 < x < \frac{\pi}{2}$.

## SOLUTIONS OF ELEMENTARY PROBLEMS

### A Double Sum Related to the Harmonic Series

E 3122 [1985, 736]. *Proposed by Miha'ly Bencze, Brasov, Romania.*

Prove that

$$\sum_{k=1}^{n} \frac{1^k + 2^k + \cdots + n^k}{kn^k}\left\{1 + (-1)^{k-1}\binom{n}{k}\right\} = (n+1)\sum_{k=1}^{n} \frac{1}{k}.$$

*Composite Solution based on the solutions of Daniel Neuenschwander (student, University of Bern, Switzerland), Bjorn Poonen (student, Harvard University), and Robert E. Shafer, Berkeley, CA.*

(student, Canada), G. Sylvester, D. B. Tyler, C. Wildhagen (The Netherlands), Yan-loi Wong (student), and the proposer.

### A Sequence of Reversals

E 3163 [1986, 566]. *Proposed by Clark Kimberling, University of Evansville, IN.*

Consider the sequence $\{\pi_i\}_{i=2}^{\infty}$ of permutations of the positive integers defined by

$$\pi_i\colon \begin{pmatrix} ki+1 & ki+2 & \cdots & ki+i \\ ki+i & ki+i-1 & \cdots & ki+1 \end{pmatrix}$$

as $k$ runs through all nonnegative integers, i.e., $\pi_i$ reverses the $i$-tuples of consecutive positive integers. Define $A_n = \pi_2\pi_3 \cdots \pi_n(1)$. Prove that the sequence

$$\{A_n\}_{n=2}^{\infty} = \{2, 4, 5, 7, 12, \dots\}$$

is strictly increasing.

*Editorial Comment.* The original statement of the problem incorrectly defined $A_n$ as $\pi_n\pi_{n-1} \cdots \pi_2(1)$. For example $\pi_3\pi_2(1) = \pi_3(2) = 2$ but $\pi_2\pi_3(1) = \pi_2(3) = 4$.

*Solution by David Callan, University of Bridgeport, CT.*

We will show that $\pi_i\pi_{i+1} \cdots \pi_n\pi_{n+1}(1) > \pi_i\pi_{i+1} \cdots \pi_n(1)$ whenever $2 \leqslant i \leqslant n$. The case $i = 2$ then gives $A_{n+1} > A_n$.

LEMMA 1. *If $b - a \equiv r \pmod{i}$ with $0 \leqslant r < i$, then $\pi_i(b) - \pi_i(a)$ is either $b - a - 2r$ or $b - a + 2i - 2r$.*

*Proof.* Expressing $b$ as $ki + j$ and $a$ as $k'i + j'$ with $1 \leqslant j, j' \leqslant i$, we have $\pi_i(b) = ki + i + 1 - j$ and $\pi_i(a) = k'i + i + 1 - j'$. Subtracting these yields $\pi_i(b) - \pi_i(a) = b - a - 2(j - j')$. Finally, $b - a \equiv r$ implies that $j - j'$ is $r$ or $r - i$.

LEMMA 2. *Suppose that $b > a$ and $b - a = qi + r$ with $0 \leqslant r < i$. If $q \geqslant r$, then $\pi_i(b) - \pi_i(a) = q'(i-1) + r'$, where $0 \leqslant r' < i - 1$ and $q' \geqslant \max\{q, r'\}$. In particular, $\pi_i(b) > \pi_i(a)$.*

*Proof.* By Lemma 1, $\pi_i(b) - \pi_i(a)$ is either $qi - r = q(i-1) + (q - r)$ or $(q+2)i - r = (q+2)(i-1) + (q + 2 - r)$; write either case as $c(i-1) + d$. Note that $c \geqslant \max(q, d)$. If $d < i - 1$ we are finished; otherwise, reduce $d$ modulo $i - 1$ to obtain $r'$, and increase $c$ accordingly to obtain $q'$.

Lemma 2 yields an inductive proof of the claim. Since $\pi_{n+1}(1) = n + 1$, take $b = n + 1$, $a = 1$, and $i = n$. Since $b - a = n$, we have $q \geqslant r$ for Lemma 2. Successive applications with decreasing $i$ yield the desired inequalities, preserving $q' \geqslant r'$ for the next application in each case.

Solved also by O. P. Lossers (The Netherlands), O. Matouš (Czechoslovakia), and the proposer.

**Infinite Integrals that are Polynomials in $\pi^2$**

E 3168 [1986, 650]. *Proposed by Douglas B. Tyler, University of California, Davis.*

(i) Show that if

$$J(a,n)=\int_0^\infty \{x(x+a)\}^{(n/2)-1}\left\{\ln\left(\frac{x}{x+a}\right)\right\}^n dx,$$

then $J(a,n)=(-1)^n a^{n-1} b_n$, where

$$b_n=\int_{-\infty}^{\infty}\left(\frac{y}{\sinh(y)}\right)^n dy.$$

(ii) Show that $b_n$ is a polynomial of degree $\lfloor (n+1)/2\rfloor$ in $\pi^2$ with rational coefficients for $n=1,2,3,\dots$ (See E 2865 [1981, 66; 1982, 426]).

*(iii) Find a closed form expression for the coefficients.

*Solution by C. Georghiou, University of Patras, Greece.* We assume throughout that $a$ is a positive real number and $n$ is a positive integer.

Set $x/(x+a)=e^{-t}$. Then

$$J(a,n)=(-1)^n a^{n-1}\int_0^\infty t^n e^{-nt/2}(1-e^{-t})^{-n}\,dt$$
$$=(-1)^n a^{n-1}\int_0^\infty \left\{\frac{t}{2\sinh(t/2)}\right\}^n dt$$

from which (i) follows.

Now

$$b_n=\int_0^\infty t^n e^{-nt/2}\sum_{k=0}^{\infty}\binom{n+k-1}{k}e^{-kt}\,dt$$
$$=n\sum_{k=0}^{\infty}\frac{(n+k-1)!}{k!}\frac{1}{(k+n/2)^{n+1}}.$$

Then we have

$$b_{2n}=2n\sum_{k=0}^{\infty}\frac{(2n+k-1)!}{k!}\frac{1}{(k+n)^{2n+1}}$$
$$=2n\sum_{k=1-n}^{\infty}(k+n)^{-2n}\prod_{i=1}^{n-1}\left\{(k+n)^2-i^2\right\}$$

since, if $1-n\leqslant k\leqslant -1$, one factor of the preceding summand is zero. Hence

$$b_{2n}=2n\sum_{j=1}^{n}(-1)^{j-1}\sigma_{j-1}\zeta(2j), \tag{1}$$

where $\sigma_0=1$, $\sigma_i$ is the $i$th elementary symmetric function of $1^2,2^2,\dots,(n-1)^2$ for $i\geqslant 1$, and $\zeta(2j)=\sum_{m=1}^{\infty}\frac{1}{m^{2j}}$. We also have

$$b_{2n-1}=(2n-1)\sum_{k=0}^{\infty}\frac{(2n+k-2)!}{k!}\frac{2^{2n}}{(2k+2n-1)^{2n}}$$
$$=4(2n-1)\sum_{k=1-n}^{\infty}(2k+2n-1)^{-2n}\prod_{i=1}^{n-1}\left\{(2k+2n-1)^2-(2i-1)^2\right\}$$

since, if $1 - n \leqslant k \leqslant -1$, one factor of the preceding summand is zero. Hence

$$b_{2n-1} = 4(2n-1)\sum_{j=1}^{n}(-1)^{j-1}\sigma'_{j-1}(1-2^{-2j})\zeta(2j) \qquad (2)$$

where now $\sigma'_0 = 1$, $\sigma'_i$ is the $i$th elementary symmetric function of the $n-1$ integers $1^2, 3^2, \ldots, (2n-3)^2$ for $i \geqslant 1$, and we have used the identity

$$\sum_{m=1}^{\infty}\frac{1}{(2m-1)^{2j}} = (1-2^{-2j})\zeta(2j).$$

Now (ii) follows from (1) and (2), since

$$\zeta(2j) = (-1)^{j-1}\frac{(2\pi)^{2j}}{2(2j)!}B_{2j}, \qquad (3)$$

where $B_{2j}$ are the Bernoulli numbers.

*Editorial Comment.* Using the Bernoulli polynomials and Newton's formulas relating elementary symmetric functions and power sums, we can express $\sigma_j$ or $\sigma'_j$ as a polynomial in $n$ of degree $2j+1$ with rational coefficients, for any particular value of $j$. For example.

$$\sigma_1 = \frac{n(n-1)(2n-1)}{6}, \qquad \sigma'_1 = \frac{(n-1)(2n-1)(2n-3)}{3}.$$

These polynomials for $\sigma_j$ and $\sigma'_j$ enable one to use (1) and (2) to compute the polynomials in $\pi^2$ mentioned in (ii). This recipe for the coefficients of the $b_n$ polynomials is effective, but we leave to the reader the decision as to whether or not it constitutes a "closed form expression."

P. L. Hon (Hong Kong), O. P. Lossers (The Netherlands), and the proposer gave somewhat similar solutions based on (3). S. J. Bang (Korea) and M. L. Glasser gave solutions based on formulas quoted from Gradshteyn and Ryzhik's *Tables of Integrals, Sums, Series, and Products* or Erdélyi, Magnus, Oberhettinger, and Tricomi's *Tables of Integral Transforms.*

## Displacement of Even and Odd Permutations

E 3175 [1986, 733]. *Proposed by C. Olmsted, University of Alaska, Fairbanks.*

Define an integer valued function $d_n$ on the set $S_n$ of permutations $\pi$ of $\{1, 2, \ldots, n\}$ by

$$d_n(\pi) = \sum_{i=1}^{n}|i - \pi(i)|.$$

It is known that the range $(d_n) = \{0, 2, 4, \ldots, \lfloor n^2/2 \rfloor\}$(cf. E 2424 [1973, 692; 1974, 668]). Let $E_n(k)$ be the number of elements in $\{\pi\colon d_n(\pi) = 2k,\ \pi \text{ is even}\}$ and $O_n(k)$ the number of elements in $\{\pi\colon d_n(\pi) = 2k,\ \pi \text{ is odd}\}$.

Determine the form of the difference $E_n(k) - O_n(k)$.

*Solution by R. H. Jeurissen, Math. Instituut, Toernooiveld, Nijmegen, The Netherlands.* Let $F_n(k) = E_n(k) - O_n(k)$, and let $G_n(k) = \{\pi \in S_n\colon d_n(\pi) = 2k\}$.

Clearly $F_n(0) = 1$ for $n \geqslant 1$, and $F_1(k) = 0$ for $k \geqslant 1$. We show that $F_n(k) = (-1)^n\binom{n-1}{k}$ for $n \geqslant 1$, $k \geqslant 0$.

Now suppose $n \geqslant 2$. The set of permutations in $G_n(k)$ for which $\pi(n-1) \neq n$ and $\pi(n) \neq n$ is mapped onto itself if its elements are preceded by the transposition $(n-1, n)$, which interchanges even and odd permutations, so these permutations contribute 0 to $F_n(k)$. The permutations in $G_n(k)$ with $\pi(n) = n$ correspond to the permutations in $G_{n-1}(k)$ by restriction, without changing parity, so these contribute $F_{n-1}(k)$ to $F_n(k)$. Finally, any permutation in $G_n(k)$ with $\pi(n-1) = n$ can be obtained by following the transposition $(n-1, n)$ with a permutation in $G_{n-1}(k-1)$ extended to leave $n$ fixed. In view of the change of parity, the permutations in $G_n(k)$ with $\pi(n-1) = n$ contribute $-F_{n-1}(k-1)$ to $F_n(k)$.

This yields the recurrence $F_n(k) = F_{n-1}(k) - F_{n-1}(k-1)$ for $n \geqslant 2$, $k \geqslant 1$, with boundary conditions $F_1(k) = \delta_{k0}$ and $F_n(0) = 1$. It follows by induction that $F_n(k) = (-1)^k\binom{n-1}{k}$.

The proposer gave a solution using generating functions. No other solutions were received.

### Highly Asymetric Graphs

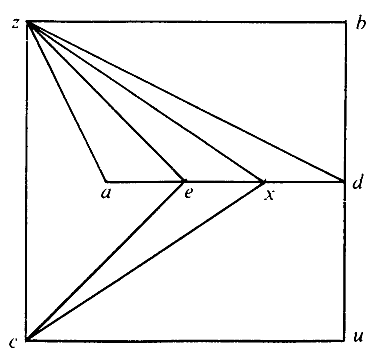
E 3187 [1987, 72]. *Proposed by Stephen J. Lipscomb, Mary Washington College, Fredericksburg, VA, and Allen J. Schwenk, Western Michigan University, Kalamazoo.*

A *vertex-deleted* subgraph $G - v$ of a graph $G$ is formed by removing one vertex $v$ and every edge incident with it. A graph is called *asymmetric* if it has no nontrivial automorphisms (symmetries).

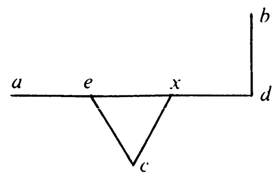
(a) Find a smallest possible asymmetric graph all of whose vertex-deleted subgraphs are also asymmetric.

(b) Same as (a) but also require that no pair of vertex-deleted subgraphs be isomorphic.

*Solution by Douglas B. West, University of Illinois, Urbana.* We interpret "smallest" to mean "fewest vertices". We show that the 8-vertex graph $F$ below has the properties of both (a) and (b) and no smaller graph satisfies either. Note that the subgraph $H = F - u - z$ is an asymmetric graph with 6 vertices and 6 edges.




F




H

Call a graph *highly asymmetric* if it is asymmetric and all its vertex-deleted subgraphs are asymmetric. We first prove $F$ is highly asymmetric. If a graph $G$ has a unique vertex $v$ of some degree, any symmetry must fix $v$, so asymmetry of $G - v$ is a sufficient (but not necessary) condition for asymmetry of $G$. Hence asymmetry of $F - u$ and $F - z$ follows from asymmetry of $H$. For any other $F - v$, $z$ is the unique vertex of $F - v$ with maximum degree and is adjacent to all but $u$, so it suffices to show that no non-trivial symmetry of $F - v - z$ fixes $u$. This can be done quickly by examination. To see that $F$ satisfies (b), note that the degree sequences of the vertex-deleted subgraphs are all distinct, except that $F - a$ and $F - b$ have a degree sequence 5444322, and $F - c$ and $F - e$ have degree sequence 5433221. In $F - b$, the vertices of degree 4 induce a triangle, while in $F - a$ they do not. Similarly, in $F - e$ the vertices of degree 1 and 5 are adjacent, while in $F - a$ they are not.

To show that no smaller graph than $F$ is highly asymmetric, we need the following facts:

Let $G$ be an asymmetric graph of smallest order. Then

(1) $G$ and $\overline{G}$ are connected.

(2) $G$ has six vertices.

(3) If $G$ has at most 6 edges, then $G = H$ above.

(4) $G$ has a vertex of degree at least 3.

(1) follows from minimality and the fact that $G$ and $\overline{G}$ have the same symmetries; any non-trivial component of a disconnected $G$ would have a non-trivial symmetry. Of $G$ and $\overline{G}$, we may consider the one with fewer edges. Thus on four vertices we need only consider trees; on five vertices we need only consider trees or connected unicyclic graphs (5 edges); every one of these graphs has a symmetry of order 2. With the existence of $H$, this proves (2). On six vertices there are six (unlabeled) trees; each has a symmetry of order 2. Among the connected unicyclic 6-vertex graphs (six edges), there are $1, 1, 4, 7$, respectively, having a cycle of length $6, 5, 4, 3$. Every one of these has a symmetry of order 2 except the graph $H$ above, which is asymmetric, proving (3). For (4), note that $\Delta(H) = 3$, where $\Delta(G)$ denotes the maximum vertex degree. The average degree of any 6-vertex graph with at least 7 edges is at least $2 \cdot 7/6 > 2$, so $\Delta(G) \geqslant 3$ even if $G \neq H$. (Note: there are also three asymmetric 6-vertex graphs with 7 edges, their complements with 8 edges, and $\overline{H}$ with 9 edges, but we need not determine the rest of the set to solve this problem.)

We now show that no highly asymmetric graph $G$ has 7 vertices. Suppose $G$ is such a graph. We first claim each vertex of $G$ has degree 2, 3, or 4. If $v$ is a vertex of degree 1, delete its neighbor. If $d(v) = 5$, delete its non-neighbor. If $d(v) = 0$ or $d(v) = 6$, delete any vertex other than $v$. In each case, we obtain a graph $G - v$ with a vertex of degree 0 or 5. This implies $G - v$ or its complement is asymmetric and disconnected, contradicting (1).

Let $w$ be a vertex of maximum degree in $G$. By (4), we have $\Delta(G - v) \geqslant 3$ for each $v$, so $d(w) \geqslant 3$. Since $G - w$ has at least 6 edges, $G$ therefore has at least 9 edges. If exactly 9, then $d(w) = 3$ and $G - w = H$, since $H$ is the only asymmetric 6-vertex graph with 6 edges. Morever, $w$ must be adjacent to $a$, $b$ and one of $\{c, d\}$. In either case, $G - a$ has a symmetry. Thus we may assume $G$ has at least 10 edges.

As before, we may assume by complementation that $G$ has no more edges than $\overline{G}$. Since $\binom{7}{2} = 21$, we may now assume $G$ has exactly 10 edges. If $\Delta(G) = 3$, then $G$

has degree sequence 3333332, and deleting the vertex of minimum degree yields the sequence 333322. The complement of this graph, with degree sequence 332222, must be asymmetric. If the two 3-valent vertices $u$, $v$ are not adjacent, then $u$, $v$ have two common neighbors, and the remaining edge joins their remaining neighbors. If $u$, $v$ are adjacent, then this degree sequence yields $2, 1, 0$ graphs when $u$, $v$ have $0, 1, 2$ common neighbors, respectively. All four graphs with this degree sequence have symmetries of order 2.

Hence we may assume $\Delta(G) = 4$, so $G - w = H$. Since $G$ has no vertex of degree 1, $w$ must be joined to $a$, $b$, and two of $\{c, d, e, x\}$. We find a symmetry in $G - c$ unless the last two neighbors of $w$ are $x$ and one of $\{d, e\}$, But then $G$ has a symmetry if $w$ is adjacent to $e$, and $G + x$ has a symmetry if $w$ is adjacent to $d$. This exhausts all possibilities for $G$.

No other solutions were received. The proposers supplied a 9-vertex highly asymmetric graph and a 10-vertex highly asymmetric graph with non-isomorphic vertex-deleted subgraphs.

### Locating Corners from an Arbitrary Point in a Rectangle

E 3208 [1987, 456]. *Proposed by I. D. Berg R. L. Bishop, and H. G. Diamond, University of Illinois at Urbana-Champaign.*

Suppose that in the Euclidean plane, line segments of lengths $a, b, c, d$ emanate from a given point $P$ in clockwise order, where $a, b, c, d$ are given positive numbers with

$$a^2 + c^2 = b^2 + d^2.$$

(i) Show that the four segments can be so placed that the endpoints determine a rectangle containing $P$, and show that this rectangle may have any specified area between 0 and some maximum value $M(a, b, c, d)$.

(ii) Find $M(a, b, c, d)$.

*Solution by Carl Schoen, University of Wisconsin at Eau Claire.* Without loss of generality, assume $a$ is the smallest of $a, b, c, d$. Contrary to the assertion of the problem as printed, the minimum area is $(a + d)\sqrt{b^2 - a^2}$ if $b \leqslant d$ and $(a + b)\sqrt{d^2 - a^2}$ if $d \leqslant b$. The area can equal zero only if $a = \min\{b, d\}$. The maximum area is $ac + bd$.

We may orient any rectangle to have horizontal and vertical sides, $P$ at the origin, and $a$ in the first quadrant. Let $\alpha$ be the angle between $a$ and the positive $x$-axis. For any choice of $\alpha$ with $0 \leqslant \alpha \leqslant \pi/2$, we construct such a rectangle. Having chosen $a$ as the smallest of $a, b, c, d$, there is a unique placement for segment $b$ so that the endpoints of $a$ and $b$ lie on a vertical line. Similarly, there is a unique placement for $d$ so that the endpoints of $a$ and $d$ lie on a horizontal line. This determines a rectangle with fourth vertex $C$. If the rectangle meets the axes at $(-t, 0)$, $(u, 0)$, $(0, w)$, and $(0, -v)$, then $b^2 + d^2 = u^2 + v^2 + t^2 + w^2 = a^2 + PC^2$, which implies $PC = c$. Hence, the postulated rectangle can be constructed for arbitrary $\alpha$.

If $A(\alpha)$ is the area of the resulting rectangle, than $A = (t + u)(v + w)$, where

$$w = a\sin\alpha, \qquad u = a\cos\alpha, \qquad t = \sqrt{d^2 - a^2\sin^2\alpha}, \qquad v = \sqrt{b^2 - a^2\cos^2\alpha}.$$

Note that $t = 0$ only if $\alpha = \dfrac{\pi}{2}$ and $d = a$, and $v = 0$ only if $\alpha = 0$ and $b = a$. Differentiating with respect to $\alpha$, we use $w' = u$, $u' = -w$, $t' = -uw/t$, $v' = uw/v$ to obtain

$$A' = (t + u)\left(\frac{uw}{v} + u\right) + \left(\frac{-uw}{t} - w\right)(v + w) = \left(\frac{u}{v} - \frac{w}{t}\right)A.$$

For $\alpha \in (0, \pi/2)$, we have $v, t, A \neq 0$, so a critical point must have $tu = vw$. This implies $a\cos\alpha\sqrt{d^2 - a^2\sin^2\alpha} = a\sin\alpha\sqrt{b^2 - a^2\cos^2\alpha}$, i.e., $\tan\alpha = d/b$.

Thus the only three candidates for extrema are $\alpha = 0$, $\alpha = \pi/2$, and $\alpha = \tan^{-1}(d/b)$. We have $A(0) = (a + d)\sqrt{b^2 - a^2}$, $A(\pi/2) = (a + b)\sqrt{d^2 - a^2}$, and $A(\tan^{-1}(d/b)) = bd + a\sqrt{b^2 + d^2 - a^2} = bd + ac$. Since $a^2 + c^2 = b^2 + d^2$, we have $A(0) < ac + bd$ and $A(\pi/2) < ac + bd$. Hence the maximum area is $ac + bd$, and the minimum area is the smaller of $A(0)$ and $A(\pi/2)$. By continuity, any area between the maximum and minimum can be obtained by suitable choice of $\alpha$.

Solved also by J. Fukuta (Japan), T. Gobinath, V. Konečný, O. P. Lossers (The Netherlands), J. G. Mauldon, V. Schindler (East Germany), T.-P. Shen (West Germany), J. H. Steelman, S. Suh, D. J. Williams & V. C. Williams, Students of 1987 Math Olympiad Program (U.S. Military Academy), and Western Maryland College Problems Group. One incorrect solution was received.

### Avoidance of Arithmetic Progressions

E 3212 [1987, 457]. *Proposed by J. Zhu, Pembroke College, Cambridge, England.*

Answer the following question mentioned by Paul Erdös in his article "Ulam, the man and the mathematician," *Journal of Graph Theory* 9 (1985), 445–449: Is it true that if $n$ is sufficiently large and $a_1, a_2, \ldots, a_n$ is any permutation of $1, 2, \ldots, n$, then there is an arithmetic progression $i, i + d, i + 2d$ with $1 \leqslant i < i + d < i + 2d \leqslant n$ such that $a_i, a_{i+d}, a_{i+2d}$ also forms an arithmetic progression?

*Solution by The University of South Alabama Problem Group, Mobile, AL.* The answer is no. For any prime $p$ define a permutation $a$ on $\{1, 2, \ldots, p - 1\}$ by the rule $a_i \equiv g^i(\text{mod } p)$, where $g$ is a primitive root modulo $p$, i.e., the residue class of $g$ modulo $p$ generates the group of units in $Z/pZ$. If $a_{i+d} - a_i = a_{i+2d} - a_{i+d}$, then $g^{i+d} - g^i \equiv g^{i+2d} - g^{i+d}(\text{mod } p)$. This implies $(g^d - 1)^2 \equiv 0(\text{mod } p)$, so that $d \equiv 0(\text{mod } p - 1)$. But now $i + d$ and $i + 2d$ necessarily exceed $p - 1$.

*Editorial Comment.* John P. Robertson points out that this is a special case of problem E 2440 [1973, 1058] for which three solutions were published [1975, 74]. The solution above is different from these.

Also solved by I. C. Bivens and L. R. King, R. B. Eggleton, C. Hurd, I. Kozma (Israel), O. P. Lossers (The Netherlands), M. Orlowski & M. Pachter (South Africa), and the proposer.

### Solutions to $\phi(x + n) = 3\phi(x)$

E 3215 [1987, 548]. *Proposed by Lászl̀ó Cseh and Imre Merényi, Cluj, Romania.*

Let $\phi$ denote Euler's arithmetical function and let $A$ be the set of positive integers $n$ for which the equation $\phi(x + n) = 3\phi(x)$ has at least one solution $x$.

Prove that for every positive integer $N$ at least half of the integers in $[1, N]$ belong to $A$.

*Solution by K. D. Wallace and R. G. Powers, Western Kentucky University, Bowling Green, KY.* We will solve the problem by showing that $A$ contains 1 and every multiple of 3 or 4. Since $\phi(7) = 3\phi(6)$, $1 \in A$. For multiples of 3, consider even and odd $n$ separately. If $n$ is odd and divisible by 3, we have $\phi(2n + n) = \phi(3n) = 3\phi(n) = 3\phi(2n)$. If $n$ is even and divisible by 3, then $n = 2k$ with $3|k$, so $\phi(k + n) = \phi(3k) = 3\phi(k)$.

It remains only to consider multiples of 4 that are not multiples of 3. Again we consider two cases, depending on whether such an $n$ is divisible by 7. If not, we have $n = 4k$ with $3 \nmid k$ and $7 \nmid k$; thus $\phi(3k + n) = \phi(7k) = 6\phi(k) = 3\phi(3k)$. If $7|n$, we have $n = 28k$ with $3 \nmid k$. Since $3 \nmid k$, we have $\phi(12k) = 2\phi(4k) = \phi(8k)$. Thus $\phi(8k + n) = \phi(36k) = 3\phi(12k) = 3\phi(8k)$.

*Editorial comment.* Note the delicacy in the solution above: if $N \equiv 2 \bmod 12$, then exactly half of the integers in $[1, N]$ have been shown to be in $A$. Solvers provided a large variety of subsets of $A$ to show its density at least .5. Both the proposers and D. B. Tyler (who stated that the density is at least .864) conjecture that the equation is solvable for all $n$ except $n = 2$. Both $\phi(x + n) = 2\phi(x)$ and $\phi(x + n) = \phi(x)$ are solvable for all $n$. (See W. Sierpinski, *Pub. Math. Debrecen* 4 (1956) 184–185 and A. Makowski, *Elem. Math.* 29 (1974) 13, respectively.

Solved also by R. B. Eggleton (Australia), A. Faccini (Italy), J. Ferrer (Spain), W. Janous (Austria), O. P. Lossers (The Netherlands), D. B. Tyler, A. Zulauf, the University of South Alabama Problem Group, and the proposer.

### Minimum Average Distance Between Points in a Rectangle

E 3217 [1987, 549]. *Proposed by Richard E. Pfiefer, San Jose State University.*

If $R$ is a plane rectangle of area 1, let $D(R)$ be the average distance between the two points of $R$ taken at random (with respect to uniform distribution). Prove that $D(R) \geqslant D(S)$, where $S$ is a square of area 1, and that equality holds only when $R$ is a square.

*Solution by A. L. Holshouser, Charlotte, NC, and L. R. King and B. G. Klein, Davidson College, Davidson, NC.* We prove more generally that the inequality holds in $n$ dimensions for a rectangular parallelepiped $R$ of volume 1 and the unit cube $C = [0,1]^n$, with equality only when $R$ is a cube.

Let $\mathbf{a} = (a_1, \ldots, a_n)$ be the side-lengths of $R$, so that $\Pi a_i = 1$, and let $D(\mathbf{a})$ also denote $D(R)$. By symmetry, If $\mathbf{b}$ is a permutation of $\mathbf{a}$, then $D(\mathbf{b}) = D(\mathbf{a})$. Now

$$D(\mathbf{a}) = \int_R \int_R \sqrt{\Sigma(u_i - x_i)^2}\, d\mathbf{u}\, d\mathbf{x}$$

where $d\mathbf{u} = du_1 \cdots du_n$ and $d\mathbf{x} = dx_1 \cdots dx_n$. Changing variables ($a_i u_i$ for $u_i$ and $a_i x_i$ for $x_i$ for all $i$) and using $\Pi a_i = 1$ normalizes the integral to the unit cube $\mathbf{C}$:

$$D(\mathbf{a}) = \int_C \int_C \sqrt{\Sigma a_i^2 (u_i - x_i)^2}\, d\mathbf{u}\, d\mathbf{x}$$

Define vectors $\mathbf{w}^1, \ldots, \mathbf{w}^n$ by $w_i^j = a_{i+j-1}(u_i - x_i)$, where subscripts are taken modulo $n$. Applying the triangle inequality to $\mathbf{w}^1 + \cdots + \mathbf{w}^n$ yields $(\Sigma a_i)|\mathbf{u} - \mathbf{x}| \leqslant \Sigma|\mathbf{w}^j|$. Integrating over choices of $\mathbf{u}, \mathbf{x}$ from $C$ and using the symmetry noted above, we find

$$(\Sigma a_i) D(C) \leqslant D(a_1, a_2, \ldots, a_n) + \cdots + D(a_n, a_1, \ldots, a_{n-1}) = nD(\mathbf{a}).$$

Thus $D(C)\Sigma a_i/n \leqslant D(R)$, with equality if $\mathbf{a} = (1, \ldots, 1)$.

We complete the proof by noting that, since $\Pi a_i = 1$, the arithmetic-geometric mean inequality gives $\Sigma a_i \geqslant n$, with equality only when $\mathbf{a} = (1, \ldots, 1)$.

*Editorial comment.* Several readers explicitly evaluated the integral of the original problem. If $D$ is the average distance between two random points in an $a$ by $1/a$ rectangle, then $D = af(a^{-2}) + a^{-1}f(a^2)$, where

$$f(t) = \frac{5t\sinh^{-1} t + (3t^2 - 2)\sqrt{t^2 + 1} + 2}{30t^2}$$

Thus $D(S) = \frac{1}{3}\ln(1 + \sqrt{2}) + \frac{1}{15}(2 + \sqrt{2}) = 0.5214\ldots$. This calculation appears as a problem in the SIAM *Review* 18 (1976), 498–499. A calculation for the 3-dimensional problem appears in this MONTHLY as problem E 2629 [1977, 57; 1978, 277].

Related results appear in *Integral Geometry and Geometric Probability* by Luis A. Santalo (Addison-Wesley, 1976) and B. Ghosh, Random distances within a rectangle and between two rectangles, *Bull. Calcutta Math. Soc.*, 43 (1951) 17–24.

Also solved by G. Bach (West Germany), G. Deridder (Canada), J. Fitch, M. S. Klamkin (Canada), jointly by A. J. Krishna (student), A. M. Rao, & G. S. Rao, L. Kuipers (Switzerland), O. P. Lossers (The Netherlands), C. A. Meyer (Switzerland), D. Neuenschwander (Switzerland), P.-O. Nyman (Finland), S. C. Pinault, P.-Y. Wu (Taiwan), Chico State Univ. Problem Group, Univ. of South Alabama Problem Group, Western Maryland College Problems Group, and the proposer. Four incorrect or incomplete solutions were received.

## ADVANCED PROBLEMS

6589. *Proposed by Kevin Brown, Kent, WA.*

Given a positive integer $g$, let $n(g)$ denote the number of conjugacy classes of Möbius transformations each member of which generates a cyclic group of order $g$. Characterize those $g$ greater than 6 for which $n(g) \equiv 1 \pmod 6$. (A Möbius transformation is a linear fractional transformation of the extended complex plane. Cf. Chapter II of Carathéodory's *Theory of Functions*, Volume I, Chelsea, 1954.)

6590. *Proposed by Jesús Ferrer, Oliva (Valencia), Spain.*

Let $Z$ be the set of integers with the topology generated by the sets $a + kZ = \{a + kx\colon\ x \in Z\}$, where $a \in Z$, $k \in Z$, $k \neq 0$. (This topology, known as the Evenly Spaced Integer Topology, is studied in considerable detail in Steen and Seebach's *Counterexamples in Topology*.)

Suppose $k$ and $l$ are relatively prime positive integers with $k > 2$, and let $B_{k,l}$ be the set of all (positive) prime numbers congruent to $l$ modulo $k$. Find the derived set of $B_{k,l}$, i.e., the set of accumulation points of $B_{k,l}$.

6591. *Proposed by Ronald J. Evans, University of California, San Diego, and Jerrold R. Griggs, University of South Carolina, Columbia.*

Let $n$ be an integer exceeding 4 and put $t = \lfloor (n-1)/2 \rfloor$. Find the eigenvalues of the following $t$ by $t$ matrices:

(i) the matrix $S$ in which the element in the $j$th row and $k$th column is $\sin(2\pi jk/n)$, $1 \leqslant j,\ k \leqslant t$.

(ii) the matrix $C$ in which the element in the $j$th row and $k$th column is $\cos(2\pi jk/n)$, $1 \leqslant j,\ k \leqslant t$.

## SOLUTIONS OF ADVANCED PROBLEMS

### An Analogue of Uniform Distribution

6542 [1987, 386]. *Proposed by Andrew M. Odlyzko, AT & T Bell Labs, Murray Hill, NJ.*

Suppose $x_1, x_2, x_3, \ldots$ is a sequence of numbers in $[0,1)$ such that at least one of its sequential limit points is irrational. For given $n$, consider the $2^n$ numbers $\varepsilon_1 x_1 + \varepsilon_2 x_2 + \cdots + \varepsilon_n x_n$, where each $\varepsilon_i$ takes the two values $\pm 1$. If $0 \leqslant a < b \leqslant 1$, let $N_n(a, b)$ be the number of $n$-tuples $(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)$ such that the fractional part of $\varepsilon_1 x_1 + \varepsilon_2 x_2 + \cdots + \varepsilon_n x_n$ is in $[a, b)$. Prove that for every $a, b$ we have

$$\lim_{n \to \infty} 2^{-n} N(a, b) = b - a.$$

*Solution by David G. Cantor, University of California at Los Angeles.* Let $f$ be a complex-valued function of period 1 defined on the real line and define

$$N_n(f) = \sum f(\varepsilon_1 x_1 + \varepsilon_2 x_2 + \cdots + \varepsilon_n x_n),$$

where the sum is over all $2^n$ of the $n$-tuples $(\varepsilon_1, \varepsilon_2,, \ldots, \varepsilon_n)$. The problem amounts to showing that

$$\lim_{n \to \infty} 2^{-n} N_n(f) = \int_0^1 f(t)\, dt \qquad (*)$$

whenever $f$ is the periodic extension of the characteristic function of the half-open interval $[a, b)$ contained in $[0,1)$.

This will certainly be true if (*) holds for all continuous periodic functions $f$ and the latter will be true if, for all integers $m$, (*) holds when $f$ is replaced by a function $f_m$ defined as $f_m(t) = \exp(2\pi imt)$ (for details of this method, due to Hermann Weyl, see Chapter IV of *An Introduction to Diophantine Approximination* by J. W. S. Cassels, Cambridge University Press, 1957, reprinted by Hafner Publishing Company, 1972).

Suppose then that $f = f_m$ is such a function. When $m = 0$, both sides of (*) are 1 and when $m$ is a nonzero integer, the right-hand side of (*) is 0. It thus suffices to prove that if $m \neq 0$, then

$$S_m(n) = 2^{-n} \sum e^{2\pi im(\varepsilon_1 x_1 + \varepsilon_2 x_2 + \cdots + \varepsilon_n x_n)} \to 0 \quad \text{as } n \to \infty,$$

where the sum is over all $2^n$ of the $n$-tuples $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$. Now,

$$S_m(n) = 2^{-n} \prod_{j=1}^{n} \sum_{\varepsilon_j = \pm 1} e^{2\pi i m \varepsilon_j x} = \prod_{j=1}^{n} \frac{e^{2\pi i m x_j} + e^{-2\pi i m x_j}}{2} = \prod_{j=1}^{n} \cos(2\pi m x_j).$$

By hypothesis, the sequence $\{2mx_j\}$ has an irrational sequential limit point, say $y$. Let $I$ be a closed interval containing $y$ and not containing any integer. Then there exists a positive $\delta < 1$ such that $|\cos \pi t| < \delta$ if $t \in I$. Let $v_m(n)$ be the number of $x_i$, $1 \leqslant i \leqslant n$ such that the fractional part of $2mx_i$ lies in $I$. Then $S_m(n) \leqslant \delta^{v_m(n)} \to 0$ as $n \to \infty$.

*Editorial Comment.* All solutions received were essentially the same as Cantor's.

Solved also by Nick Martin, A. Meir (Canada), James Propp & Daniel Ullman (jointly), Arthur Rothstein, and the proposer.

---

## What Else Pythagoras Could Have Done

EDITOR:

Yoram Sagher [See reference] shows that Pythagoras could have demonstrated that the square root of an integer is an irrational or an integer without knowing anything about prime numbers. Assuming Pythagoras understood Euclid's algorithm, the following proofs show how he could have demonstrated that any integer root of an integer is an irrational or an integer, and even that the cube root of an integer either is not the root of a quadratic (i.e., not of the form $(a + b\sqrt{n})/c$) or is an integer.

In the following, all variables but $r$ are restricted to integer values.

The (shorter) proof of Sagher's special case comes first, to motivate the others. If $r = \sqrt{k} = m/n$ (in lowest terms), Euclid's algorithm gives $\alpha$ and $\beta$ for which $\alpha m + \beta n = 1$. Then $rn = m$, $rm = r^2 n = kn$, and $r = r(\alpha m + \beta n) = \alpha rm + \beta rn = \alpha kn + \beta m$, an integer.

Now take $r = m/n = \sqrt[3]{k}$, $\alpha m + \beta n = 1$. Then

$$m = rn, \qquad rn^2 = mn, \qquad rmn = m^2, \qquad rm^2 = r^3 n^2 = kn^2$$

$$r = r(\alpha m + \beta n)^2 = \alpha^2 rm^2 + 2\alpha\beta rmn + \beta^2 rn^2 = \alpha^2 kn^2 + 2\alpha\beta m^2 + \beta^2 mn,$$

an integer.

By an obvious generalization, for *any integer* $t \geqslant 2$, if $\sqrt[t]{k}$ is rational, it is an integer.

Now make the weaker assumption that $r = \sqrt[3]{k}$, and that $r$ satisfies a proper quadratic equation $ar^2 + br + c = 0$, $a \neq 0$. Then $0 = (ar^2 + br + c)(ar - b) = a^2 r^3 + (ac - b^2)r - bc$. If $ac = b^2$, divide the equation by $a^2$ to find $r^3 = (b/a)^3$ and $r = b/a$. Otherwise, put $k$ for $r^3$ and find $r = (bc - a^2 k/(ac - b^2)$. Either way, $r$ is rational, and consequently an integer.

Reference

Y. Sagher, What Pythagoras Could Have Done, *American Mathematical Monthly*, 95 (1988) 117.

ROBERT W. FLOYD

# REVIEWS

EDITED BY JOSEPH KONHAUSER
*Department of Mathematics, Macalester College, St. Paul, MN 55105*

*Mathematical Problem Solving.* By Alan H. Schoenfeld. Academic Press, Inc., Orlando, 1985. xvi + 409 pp.

STEVEN GALOVICH
*Department of Mathematics, Carleton College, Northfield, MN 55057*

The publication of George Pólya's *How to Solve It* in 1945 initiated a new era in mathematics education. This era began quite modestly, but by the late '50s, mathematicians and mathematics educators, many of them influenced by Pólya's ideas, were running summer institutes for secondary teachers and were advocating a revamping of the mathematics curriculum to include the teaching of problem solving and concepts of modern mathematics. The outcome of these efforts was, of course, the famous (or infamous) "New Math." Although it is unclear exactly how many mathematics students were touched by SMSG and its derivatives, the new math movement had a considerable impact upon American education, not the least of which was the "back-to-basics" backlash which arose in the '70s. Partly in response to the perceived failures of the new math movement and partly as a reaction to the back-to-basics movement, the mathematics education community returned to Pólya in the late '70s. The 1980 yearbook of the National Council of Teachers of Mathematics was devoted to problem solving. By declaring the '80s to be "The Decade of Problem Solving," the NCTM and the mathematics education community have, in effect, put all of their pedagogical eggs into one basket. As we near the end of the Problem-Solving Decade, it is only reasonable to assess the results of this stress on problem solving in the teaching of mathematics.

How successful has the effort to redirect secondary mathematics education toward problem solving been? Many of us who teach freshman-level mathematics courses in college would say politely: "Not very." Most of us probably feel that students entering college today are not better prepared, or perhaps are even less prepared, than those entering college ten years ago. Of course, in all fairness, the Problem-Solving Decade is not over and we should not judge prematurely these important and ambitious efforts to change the course of mathematics education in this country. In any case, such seat-of-the-pants judgments are, if anything, comments on the general state of American mathematics education and not necessarily indications of the failure of the teaching of problem solving.

On the other hand, several of us have tried to teach problem solving à la Pólya in our various mathematics courses. We pass out the famous chart from *How to Solve It*, delineating Pólya's four phases of problem solving—Understand the Problem, Devise a Plan, Carry Out the Plan, Look Back. We plow through the syllabus of the course, while trying to weave in Pólya's teachings on heuristics. And in the end, we find that our efforts are, at best, moderately successful. Moreover, one can dig into the literature on mathematics education to cite evidence which shows that the teaching of heuristics does not necessarily improve students' ability to solve problems. Thus we seem to face the unhappy situation of wanting to do something to

help our students become better problem solvers, taking what seem to be the obvious steps to bring about this goal, and discovering that our efforts are only marginally successful. Can this lack of success be explained? Is there anything that we as individuals, or as mathematicians collectively, can do to improve upon this situation?

My own feeling, stated without exaggeration, is that any attempt to answer these questions must begin with a reading of *Mathematical Problem Solving* by Alan Schoenfeld. Not only does Schoenfeld offer sound reasons for why attempts to teach problem solving often come up short of the mark, but he also postulates an interesting theory of mathematical behavior which is useful in analyzing a student's work on a particular problem and which can influence the way all of us teach mathematics. What is Schoenfeld's theory of mathematical behavior?

Schoenfeld claims that an individual's mathematical behavior is determined by four factors: (1) the person's mathematical *resources*; (2) his ability to use *heuristics*; (3) his skill at *controlling* his use of resources and heuristics; and (4) his *beliefs* about mathematics and about himself. Thus, success or failure at a mathematical endeavor is determined by Resources, Heuristics, Control, and Beliefs. Exactly what do each of these categories entail?

One's resources in a particular subject consist of all the theorems, definitions, algorithms, and proofs of theorems relevant to that particular subject. In addition, one's resources include one's understanding of the rules of logic and the standard mathematical proof techniques. Heuristic principles are the rules of thumb that one employs when attempting to solve mathematical problems. Heuristics constitutes the subject matter of Pólya's *How to Solve It*. Control refers to the students' ability to manage and to employ his resources and his knowledge of heuristics. A student who exhibits good control will, for instance, stick to the problem at hand and tend not to go off on a wild goose chase in search of a solution. Finally, one's belief systems consist of attitudes about oneself and about mathematics which affect one's ability to do mathematics. Such attitudes might concern the individual problem solver: "I never could do mathematical proofs." They might concern the doing of mathematics itself: "If you can't solve a problem within five minutes, then you might as well give up because the problem is impossible." Or they might reflect an individual's conception of the nature of mathematics: "The theorems of formal mathematics that are proved in class are of little use when trying to solve homework problems." (More on this belief shortly.)

Schoenfeld argues that a breakdown or deficiency in any one of these four areas can torpedo an individual's effort to solve a mathematical problem or to learn mathematics. For example, as most of us would probably agree, a student who is reasonably familiar with the definitions and theorems of a particular subject and yet who has a poor sense of heuristics would have a great deal of trouble solving problems. Schoenfeld cites examples of students who have a mastery of the resources of a given topic and a good command of heuristics and who, nonetheless, fail to solve problems well within their grasp because they mindlessly follow blind alleys and do not stand back to assess their progress at crucial points in the problem-solving venture. In other words, failure at the control level can impede progress toward a solution to a problem. Finally, in one of the most interesting and intriguing portions of the book, Schoenfeld argues that students possessing beliefs that are, in effect, antimathematical often fail to solve problems in spite of their

having the appropriate resources and heuristics at their disposal and having adequate control over these resources and heuristics. Let us consider this last point in more detail.

In order to analyze how students solve mathematical problems, Schoenfeld video-taped pairs of college freshmen as they worked on the following "non-standard" problem from high school geometry: Given two intersecting straight lines, and a point $P$ on only one of the lines, show how to construct, using straightedge and compass, a circle that is tangent to both lines and has the point $P$ as its point of tangency to one of the lines. Nearly all the pairs of students in Schoenfeld's admittedly small sample approached this problem in an *empirical* way. In most cases, the students video-taped by Schoenfeld justified or rejected a potential solution to the problem via pictures. In no instance did students use theorems of Euclidean geometry as investigative tools or as the basis upon which an argument was judged. The students simply did not believe that the theorems and constructions of formal geometry could be of any use in solving this problem. Although many of these students independently demonstrated knowledge of all the theorems and constructions relevant to the problem, they could not apply the pertinent results simply because, Schoenfeld claims, they have never learned that the theorems of proofs presented formally in the classroom can and should be used when trying to solve problems of this sort.

Let's return to the question as to why attempts to teach heuristics often fall short. The general answer probably comes as a surprise to no one: The teaching of heuristics fails to meet expectations because heuristics itself is a complicated subject. Specifically, Schoenfeld points out that many of the heuristic strategies proposed by Pólya are quite broad and vague and therefore difficult to implement in a given case. Such words of advice as "find a related problem" and "aim for subgoals" or "try special cases" can have a variety of meanings, depending on the given problem. Secondly, one cannot use heuristic strategies blindly. One must monitor one's use of heuristics carefully lest one become mired in a maze of possible approaches to a given problem. Thus the issues of heuristics and control are related. Finally, Schoenfeld notes that heuristics and resources are closely connected. Thus, for example, a student with a weak grasp of the resources of a particular area of mathematics will have little success in trying to apply general heuristic strategies to that domain. Moreover, that student will not understand any heuristic strategy that is specific to that domain.

Schoenfeld has made a significant contribution to our understanding of the learning and teaching of mathematics. His four-part division of mathematical behavior—Resources, Heuristics, Control, and Beliefs—provides a useful framework which we can use as a guide in developing our own philosophy and practice of teaching and as a gauge against which to measure our students' progress. The book contains much else of value, including an interesting discussion of the methodological issues that arise when trying to determine students' mathematical behavior, excellent summaries of work in cognitive psychology and artificial intelligence on problem solving, and descriptions of apparently successful attempts to teach heuristics and control to college students. All this is not to say that the entire book should be taken as gospel. For example, Schoenfeld's experiments on students' problem-solving behavior and on the teaching of problem solving should be reproduced and, if necessary, reinterpreted. In general, further investigations of the processes by

which young adults learn mathematics and solve mathematical problems must be undertaken. Perhaps out of such research will arise other useful paradigms of mathematical behavior. For the time being, this book offers to teachers of mathematics an honest and thorough analysis of the complex issue of mathematical problem solving by humans. Schoenfeld is neither optimistic nor pessimistic about the possibility of teaching heuristics and control successfully to students. It is apparent from the pedagogical experiences which Schoenfeld describes that some students can learn to use heuristic strategies wisely. It is equally apparent that the teaching of heuristics requires considerable thought and careful planning on the part of the instructor.

Should heuristics be taught in mathematics courses? Any answer to this question is certainly debatable. Some mathematicians firmly believe that heuristics cannot, in fact, be taught, and that students learn how to solve problems only by solving or trying to solve problems and by assimilating the resulting experiences into their mathematical repertoire. Those who believe otherwise can take comfort in Schoenfeld's success in teaching heuristics. Schoenfeld's examples (described in Chapter 7) were, however, carried out in small seminars and probably with fairly good students. Can heuristics be taught to the masses in a calculus course or some other mathematics course? The answer to this question is unclear. But for researchers in the area of mathematical cognition and for teachers of mathematics who believe in the teaching of heuristics, the question offers a tantalizing challenge. Schoenfeld's book describes the many factors which we must consider when we try to teach heuristics. It remains for us to assess Schoenfeld's analysis of mathematical problem solving, to test his ideas about mathematical behavior, and, by accepting, rejecting or modifying his framework, to find ways of teaching heuristics effectively.

It is customary for a reviewer who is especially enthusiastic about the book under review to insist that his audience read the book, the entire book, immediately. It might, however, be unrealistic to expect mathematicians to plow through a 400-page book and especially those passages dealing with methodological issues. Certainly anyone doing research in mathematics education or directly involved in the teaching of teachers should read all ten chapters of the book. Teachers of mathematics at both the secondary and the collegiate level would do well to read the first five chapters in which Schoenfeld's framework is discussed in detail. If you are really strapped for time, then just read Chapter 1 in which the book's thesis is outlined and the major arguments are presented.

*Geometric Theory of Foliations.* By Cesar Camacho and Alcides Lins Neto. Birkhäuser Boston, Inc. 1986. 205 pp.

Anthony Phillips
*Department of Mathematics, State University of New York, Stony Brook, NY 11794*

The geometric theory of foliations must be entering its mature phase, since after many years of no books at all there have recently been published no fewer than three, this one and [**3**, **5**] (Bruce Reinhart's book [**10**] is about the *differential* geometry of foliations, another matter entirely). Each of these books has its

attractive features, but certainly the most ambitious, polished and complete is the work considered here.

This review begins with an informal look at foliation theory, especially at the geometric side. This should serve as background for the examination of the book, and also, I hope, as propaganda to interest the general reader in this beautiful area of mathematics. Bibliographical references for works before 1977 can be found in one or the other of Blaine Lawson's surveys **[8, 9]**.

The kind of space we will foliate is a smooth $n$-dimensional manifold, i.e., a topological space locally homeomorphic to $R^n$ (*smooth* means that the coordinate changes involved in going from one of these "local charts" to another are supposed to be infinitely differentiable). To foliate such a space means to fill it completely with roughly parallel manifolds ("leaves") of some lower dimension $k$; such a foliation is said to have dimension $k$ and codimension $n - k$. By "roughly parallel" I mean that one can choose coordinates $x_1, \ldots, x_n$ in a neighborhood about any point such that in those coordinates the leaves through that neighborhood appear as the $k$-planes (rigorously) parallel to a fixed $k$-plane in $R^n$. So, for example, the collection of lines in $R^2$ given by the $x$-axis, vertical lines filling up the open upper half-plane, and vertical lines filling up the lower half-plane (FIG. 1a) is *not* a foliation, since no point on the $x$-axis has such a neighborhood.

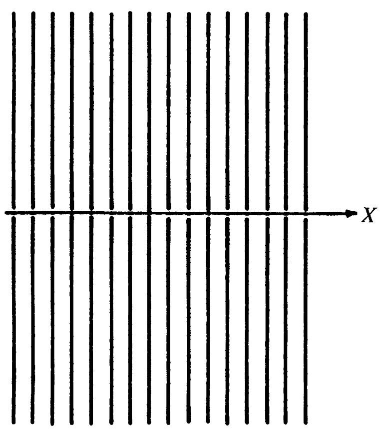



FIG. 1a.

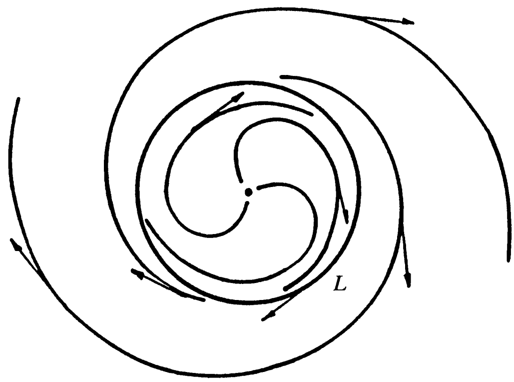



FIG. 1b.

The foliation is called of class $C^r$ if the coordinate functions mentioned above can be chosen to be $r$ times differentiable.

Example 1. *Dynamical systems*. Let $V$ be a nonzero tangent vector field defined on an open subset $\Omega$ of a smooth manifold. It is a classical result **[11]** that the integral curves of $V$ are the leaves of a 1-dimensional foliation as described above. An *integral curve* of $V$ is the orbit described by a particle moving in $\Omega$ so that when it is at point $x$ its velocity vector is $V(x)$. FIG. 1b shows a dynamical system on the punctured plane $R^2 - \{0\}$. There is one compact leaf $L$, a circle. The orbits inside $L$ spiral out towards $L$ as $t \to \infty$, while those outside spiral away from $L$ towards infinity.

The dynamical description of the leaves in this example is important in our context because, as I understand it, the geometric theory of foliations is precisely the study of foliations *as dynamical systems*. The asymptotic behavior of the orbits is enriched by interaction with the topology of the leaves. For example, an open cylinder can appear in a 2-dimensional foliation as a *resilient leaf* (see below), a type of asymptotic behavior impossible in the 1-dimensional case.

The $k$-dimensional analogue of a vector field (strictly speaking, of the underlying tangent line field) is a field of tangent $k$-planes, but the corresponding $k$-dimensional integral submanifolds only exist if the field is *integrable*. Integrability is an infinitesimal property; the criterion is given by the following classical theorem (e.g., **[11]**).

FROBENIUS' THEOREM. *A field L of k-planes is integrable iff whenever X and Y are vector fields contained in L, their Lie bracket [X, Y] is also contained in L.*

Example 2. *The Reeb foliation*. This is a 2-dimensional foliation of the 3-sphere $S^3$. Take $S^3$ (the set of points in $R^4$ satisfying $x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1$) as the union of two solid tori: $\{x_1^2 + x_2^2 \leqslant x_3^2 + x_4^2\}$ and $\{x_1^2 + x_2^2 \geqslant x_3^2 + x_4^2\}$ with common boundary the torus $\{x_1^2 + x_2^2 = x_3^2 + x_4^2 = 1/2\}$. Now foliate each solid torus so that the common boundary is a leaf. (The other leaves, diffeomorphic to $R^2$, have the shape of elongated paraboloids wrapped around so as to engulf themselves over and over again, as in FIG. 2.) If this is done smoothly enough, it gives a $C^\infty$-foliation of $S^3$.

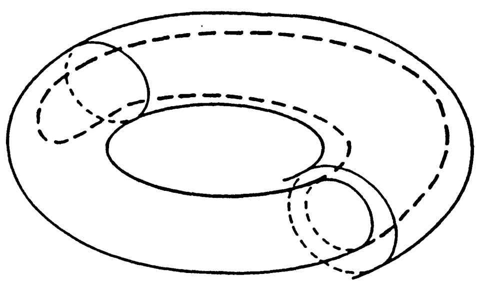


FIG. 2

The modern history of foliations began in Strasbourg in 1948, with Georges Reeb's description of this foliation in his thesis (just as the history of fiber bundles and of modern homotopy theory began with Heinz Hopf's fibration of the 3-sphere in 1931). In fact, the early period of this history all took place in a fairly small radius about Strasbourg. The dominant figures, besides Reeb, were Charles Ehresmann (Reeb's advisor) and André Haefliger. Haefliger was Georges de Rham's student in Lausanne, but any paragraph of his thesis identifies it instantly as a product of the Ehresmann school. This work, published in 1958, is remarkable because, besides being fundamental in the geometric theory of foliations, it defined and studied more general objects called "$\Gamma_q$-structures," which turned out, 10 years later, to be the natural context for the quantitative theory which was just then being developed.

Haefliger's geometric results are based on the very important concept of *holonomy*, which was first defined, as far as I can tell, in a 1956 paper by Ehresmann

and Shih Wei-Shu, although the idea is implicit in Reeb's thesis. The object is to have a precise concept of how leaf $A$ can "spiral in" on leaf $B$, in a $k$-dimensional foliation of an $n$-manifold. It is pretty clear that interesting behavior can only occur if $B$ is not simply connected; one therefore watches closed loops in $B$. Let $l$ be such a loop, and let $T$ be a piece of $n-k$-dimensional surface transverse to $B$ at $x$. Now imagine sliding $T$ along $l$, keeping each point of $T$ in the leaf in which it started. Some neighborhood of $x$ in $T$ will wind up mapped diffeomorphically into $T$: this is a holonomy diffeomorphism.

The holonomy in the Reeb foliation, as is easy to check, is very simple. If a transversal is slid around a latitude circle on the central torus, points on the "outside" come back to themselves, whereas points on the "inside" spiral in towards the torus; and vice-versa for a longitude. But there are other possibilities. A 2-dimensional leaf in a 3-manifold can sometimes twist around so as to spiral in *on itself*. This is a *resilient leaf* [**7**]. The phenomenon sounds improbable but is actually quite common; in the well-known horocycle foliations [**1**, **10**] for example, almost every leaf is resilient.

In 1965 Sergei Novikov published the following theorem.

THEOREM. *Every class-$C^2$, 2-dimensional foliation of the 3-sphere has a compact, torus leaf.*

The intricate proof required a vigorous and ingenious use of Haefliger's and Ehresmann's ideas. Although this theorem had few consequences and proved impossible to generalize, it had a profound psychological impact; it suddenly became clear to topologists in general that here was a large mathematical continent awaiting exploration.

The list of references in Blaine Lawson's survey article of May, 1973 [**8**] documents the ensuing explosion in research. Much of this was due to the new "quantitative" theory [**9**] which sprang up around 1970. Although this will take us away for a moment from our main topic, I would like to sketch out at least the beginnings of the new theory because quite recently the geometric and quantitative lines of research have been brought together in a very neat way.

The suspicion that some $p$-plane fields could not even be deformed into integrable ones had already been raised in Reeb's thesis, but the first hard evidence came from Raoul Bott, with his Vanishing Theorem, proved in 1968. He showed that some algebraic-topological invariants of a $p$-plane field, thought of as a $p$-dimensional vector bundle, had to be zero for it to be "integrable-able."

Shortly after this discovery came the Godbillon-Vey invariant. Claude Godbillon spoke at the foliations meeting in Oberwolfach in May, 1971, explaining that if on a compact, orientable, 3-manifold $M$, a codimension-one foliation $\mathbf{F}$ was defined by the kernel of a 1-form $\theta$ with (dual version of Frobenius' theorem) $d\theta = \lambda \wedge \theta$, then the closed form $\lambda \wedge d\lambda$ gave, when integrated over $M$, a real number depending only on $\mathbf{F}$. He went on to say that it was not known if this number could be nonzero. This state of affairs did not last very long. Robert Roussarie was at that meeting and shortly afterwards came out (as did William Thurston) with the observation that the horocycle foliations had nontrivial "Godbillon-Vey invariant."

These two discoveries (which were found to be closely related) formed the bridge over which the hordes of Homotopy Theory swarmed onto our unexplored continent. In addition, the assignment $X \to \{$homotopy classes of $\Gamma_q$-structures on $X\}$

turned out to be a homotopy functor and therefore to have a classifying space, etc. There was no stopping them.

With this new wave of activity came the discovery of many new foliations. In 1971 Lawson produced a codimension-one, class-$C^\infty$ foliation of $S^5$ (and of any sphere of dimension $2^k + 3$); soon all odd spheres were foliated, and not too much later Thurston showed that on a compact manifold, any codimension-one field of planes could be deformed to an integrable one. Thurston's foliations are locally very complicated; he gives as a problem (it remains outstanding in general) to replace them by foliations with the same degree of intelligibility as Reeb's and Lawson's. For results in higher codimensions, see [**4**, **12**].

One topic which has continued to interest the geometric foliation theorists is the topology and asymptotic behavior of noncompact leaves in compact, foliated manifolds. A recent, striking result is the theorem I hinted about earlier, which unites the quantitative and geometric aspects of foliations in such a satisfying way. It is not a coincidence after all that the horocycle foliations have both resilient leaves and nonvanishing Godbillon-Vey invariant.

THEOREM (Gérard Duminy; unpublished. For proof, see [**2**] and [**6**].). *If a class-$C^2$, codimension-one foliation of a compact 3-manifold has no resilient leaves, it must have Godbillon-Vey invariant zero.*

Thurston remarked that the local nonvanishing of the Godbillon-Vey form $\lambda \wedge d\lambda$ could be interpreted geometrically as a "helical wobble" of the tangent planes to the foliation; so Duminy's theorem means that if the average helical wobble is nonzero, it must manifest itself in resilient leaves.

Camacho and Lins Neto's book is the first comprehensive introduction to this field. A capsule summary of the book would state that it contains the proof of Novikov's compact leaf theorem and all the material necessary to understand it (except for the Poincaré–Bendixson Theorem; this is, however, readily available elsewhere). We should all be grateful to the authors for their efforts. Aside from Haefliger's 1967–68 Bourbaki seminar report, this is the first expository account of Novikov's theorem (another is promised for part C of [**5**]). Also, there is an interesting chapter on foliations coming from group actions. Along the way several important examples are carefully worked out and presented in detail; for example, Denjoy's $C^1$-foliation of $T^2$ and Richard Sacksteder's $C^2$-foliation of $M \times S^1$ ($M$ a surface of genus 2), each with an exceptional minimal set. This rich set of significant examples, highlighted in addition by many fine illustrations, is the great strength of the book, and should make it an indispensable asset for any student of the field.

Unfortunately, when we pass from examples to definitions, theorems and proofs, the going gets much rougher. I find it surprising, given the immense amount of effort that obviously went into the production of this book, that a little more care was not taken in the final stages of preparation of the manuscript. The publishers do not seem to have had it read through carefully to eliminate redundancies, inconsistencies in terminology, references to generally unavailable textbooks, or to the wrong article, mislabelled figures, mistranslations and, of course, misprints. In addition, the index is grossly insufficient and the references, themselves infested by a multilingual circus of misprints, are listed one after the other, in no particular order whatsoever. Word processors were designed to take care of precisely this type of mess.

Most of these blemishes are more insulting than injurious; let me just point out one of the instances where careless writing or editing may end up perplexing the ideal graduate student to whom this book is addressed.

What is a foliation of class $C^r$? In this book (p. 22):

"*Definition*. Let $M$ be a $C^\infty$ manifold of dimension $m$. A $C^r$ foliation of dimension $n$ on $M$ is a $C^r$ atlas **F** on $M$ which is maximal with the following properties... ."

Looking back to p. 7 for the definition of "$C^r$ atlas" we find the usual one: a collection of local charts that are $C^r$-interrelated on overlaps. Only lower on p. 22 do we find

"*Remark* 1. When we say $M$ is a $C^\infty$ manifold which has an atlas **F** as above, we are implicitly saying that $M$ has an atlas **A** whose changes of coordinates are $C^\infty$; however if $(U, \phi) \in \mathbf{A}$ and $(V, \psi) \in \mathbf{F}$ and $U \cap V \neq \varnothing$ then $\phi \circ \psi^{-1}$ and $\psi \circ \phi^{-1}$ are $C^r$. The only relation between **F** and **A** is that the mixed changes of variables, as above, are $C^r$ (...)."

Notice the phrase "we are implicitly saying" in Remark 1. There is nothing *implicit* in the definition. It says what it says. What they really mean is "we really mean."

REFERENCES

1. V. Arnol'd and A. Avez, Ergodic Problems of Classical Mechanics, Benjamin, New York, 1969.
2. J. Cantwell and L. Conlon, The dynamics of open, foliated manifolds and a vanishing theorem for the Godbillon-Vey class, *Advances in Math.*, 53 (1984) 1–27.
3. C. Godbillon, Feuilletages, Etudes géometriques I, Institut de Recherche Mathématique Avancée, Strasbourg, 1985.
4. A. Haefliger, Feuilletages sur les variétés ouvertes, *Topology*, 9 (1970) 183–194.
5. G. Hector and U. Hirsch, Introduction to the Geometry of Foliations, Vieweg, Braunschweig, Part A, 1981; Part B, 1983.
6. S. Hurder, The Godbillon measure of amenable foliations, *J. Diff. Geom.*, 23 (1986) 347–365.
7. C. Lamoureux, Sur quelques phénomènes de captage, *Ann. Inst. Fourier*, 23 (1973) 229–243.
8. H. B. Lawson, Jr., Foliations, *Bull. Amer. Math. Soc.*, 80 (1974) 369–418.
9. ______, The Quantitative Theory of Foliations (Regional Conference Series in Math.; 27), Amer. Math. Soc., 1977.
10. B. Reinhart, Differential Geometry of Foliations: the Fundamental Integrability Problem, Springer, Berlin; New York, 1983.
11. M. Spivak, A Comprehensive Introduction to Differential Geometry I (2nd ed.), Publish or Perish, Berkeley, 1979.
12. W. Thurston, The theory of foliations of codimension greater than one, *Comment. Math. Helv.*, 49 (1974) 214–231.
13. W. Thurston, Existence of codimension-one foliations, *Ann. of Math.* (2) 104 (1976) 249–268.

# TELEGRAPHIC REVIEWS

*Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.*

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

| | | |
|---|---|---|
| T: Textbook | P: Professional Reading | 1–4: Semesters |
| C: Computer Software | L: Undergraduate Library | **: Special Emphasis |
| S: Supplementary Reading | 13: Grade Level | ??: Questionable |

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

**General, S, L.** *Return to Mathematical Circles.* Howard W. Eves. PWS-Kent, 1987, xxi + 181 pp, $26.50. [ISBN: 0-87150-105-8] The fifth volume of *Mathematical Circles*—360 degrees of puns, anecdotes, trivia, problems, rumors, coincidences, and aphorisms, all with some connection to mathematics. Includes nearly a full quadrant of Einstein stories, and an epilogue of minutes on humor, on teaching, and on logic. LAS

**General, P, L*.** *Mathematical Sciences: Some Research Trends.* Board on Mathematical Sciences. National Academy Pr, 1988, viii + 126 pp, (P). A synopsis of selected highlights of recent research in the mathematical sciences—in applied mathematics, in core mathematics, and in statistical sciences. Three primary overview essays are supplemented by eight individually-authored vignettes providing in-depth examples ranging from soliton theory to resampling methods. High-quality exposition at a high level of sophistication: a superb source for mathematicians to learn what their colleague's research is all about. LAS

**General, P, L**.** *Michael Atiyah: Collected Works, V. 1-5.* Michael Atiyah. Clarendon Pr, 1988. *Volume 1: Early Papers: General Papers,* xxiii + 364 pp, $59 [ISBN: 0-19-853275-X]; *Volume 2: K-Theory,* xxiii + 829 pp, $89 [ISBN: 0-19-853276-8]; *Volume 3: Index Theory: 1,* xxiii + 593 pp, $79 [ISBN: 0-19-853277-6]; *Volume 4: Index Theory: 2,* xxiii + 617 pp, $79 [ISBN: 0-19-853278-4]; *Volume 5: Gauge Theories,* xxiii + 685 pp, $89. [ISBN: 0-19-853279-2] A monument of scholarship recording the accomplishment of one of the greatest living mathematicians. Each section is introduced by a commentary prepared by the author to document the genesis of ideas. No biographical essay, but does include an interview that gives rich, autobiographical insights. LAS

**Elementary, T**, L.** *Geometry: A High School Course, Second Edition.* Serge Lang, Gene Murrow. Springer-Verlag, 1988, xii + 394 pp, $39. [ISBN: 0-387-96654-4] A text which may help revitalize high school geometry; this hardcover edition incorporates a few changes from the preliminary *First Edition* (TR, January 1984; Extended Review, April 1986), e.g., new illustrations and new sections on the volume of a ball, the surface of a sphere, and isometries as compositions of reflections. JNC

**Mathematics Appreciation, T(14-15: 1), S*.** *Alice in Numberland: A Students' Guide to the Enjoyment of Higher Mathematics.* John Baylis, Rod Haggarty. Macmillan Education, 1988, ix + 205 pp, £8.95 (P). [ISBN: 0-333-44242-3] Entertaining introduction to mathematical logic and proofs. Good supplement for first "proof course." Topics include symbolic logic, unique factorization, several versions of induction, permutations, axiomatic and constructive approaches to the real numbers, sequences, series. SB

**Precalculus, T(13: 1).** *Trigonometry.* Jerome E. Kaufmann. Ser. in Math. Prindle, Weber, & Schmidt, 1988, xii + 293 pp, $27. [ISBN: 0-534-92106-X] "Should the trigonometric functions be introduced as functions of angles or functions of real numbers? This is probably the number one issue relative to the teaching of trigonometry." This text opts for angles first, circular functions later. Standard topics; three types of problems (regular, miscellaneous, review); chapter summaries; and selected calculator applications. LAS

**Precalculus, T*(13: 1).** *College Algebra and Trigonometry, Second Edition.* John R. Durbin. Wiley, 1988, xii + 593 pp, $40.88. [ISBN: 0-471-62545-0] Covering standard topics of college algebra and trigonometry, this edition includes a more streamlined treatment of the review material in the opening chapter, an expanded treatment of conic sections, more calculator exercises, a rewriting of the section on exponential and logarithmic functions, and sample assignments. Lots of examples and exercises; clearly written. (*First Edition,* TR, December 1984.) CEC

**Finite Mathematics, T(13-14: 1).** *Finite Mathematics, Second Edition.* Karl J. Smith. Brooks-Cole, 1988, xiii + 433 pp, $30. [ISBN: 0-534-08904-6] For a one-semester course in non-calculus mathematics for students in business, management, and life or social science. Prerequisite is intermediate algebra. Readable; lots of worked-out examples and drill-type problems as well as problems featuring applications and modelling. Appendices on logic and mathematical induction. Student's supplement, instructor's supplement, computerised test bank, and computer supplement are available. Author's experience as a writer of textbooks makes this book an attractive candidate for adoption. (1975 Scott Foresman edition, TR, June-July 1975.) JK

**Finite Mathematics, T(13: 2).** *College Mathematics, Second Edition.* S.T. Tan. PWS-Kent, 1988, xvi + 1147 pp, $31.50. [ISBN: 0-534-91791-7] Changes in this edition include a more traditional approach to systems of equations, reordering of a couple of topics, and the addition of a precalculus review, differentials, applications to probability, numerical integration, and the least squares method. (*First Edition,* TR, August-September 1983.) JNC

**Finite Mathematics, T(13).** *Finite Mathematics with Applications for Business and Social Sciences, Fifth Edition.* Abe Mizrahi, Michael Sullivan. Wiley, 1988, xviii + 654 pp, $45.14. [ISBN: 0-471-85291-0] Very similar to *Fourth Edition.* Three main divisions are linear algebra—aimed at elementary explanation of the simplex method (5 chapters); probability—with introductions to statistics, Markov chains (6 chapters); and discrete mathematics—logic, sequences and induction, graphs (3 chapters). (*First Edition,* TR, October 1973; *Third Edition,* TR, June-July 1979; *Fourth Edition,* TR, February 1984.) AWR

**Education, P*.** *Keep Up with Teaching Mathematics in France.* IGR Imprimerie Lyon, 1988, 86 pp, (P). [ISBN: 2-902-680-47-3] A survey of recent history and present objectives of secondary-level mathematics education in France. Central chapters outline curriculum objectives in mathematics for (essentially) all French students who finish the "Second" level (age 15-16), including numerical activities, statistics, functions, geometry, scalar product, and systems of linear equations. The teacher's ideal "should not be to offer pupils a too-perfect account; his primary task is to lead his pupils, by means of situations taken in their natural complexity, to reflection and personal autonomy." LAS

**Education, L.** *The Mathematics Teacher: Cumulative Index, Volumes 69-78, 1976-1985.* NCTM, 1988, 59 pp, $10 (P). [ISBN: 0-87353-262-7] Author and subject indices to a recent ten-year chunk of *The Mathematics Teacher.* LAS

**Education, P, L.** *Number Concepts and Operations in the Middle Grades, Volume 2.* Ed: James Hiebert, Merlyn Behr. Res. Agenda for Math. Educ. NCTM, 1988, x + 270 pp, (P). [ISBN: 0-87353-265-1] Thirteen major papers setting forth current knowledge about how number concepts mature in the middle school years, together with analyses of the effects of instruction: "It is clear that conventional instruction is inadequate in some very fundamental ways." Part of NCTM's four-volume contemporary agenda for research in mathematics education. (*Volume 1,* TR, June-July 1988.) LAS

**Education, S, P.** *Projects to Enrich School Mathematics.* Ed: Leroy Sachs. NCTM, 1988. *Level 2,* ii + 93 pp, $6 (P) [ISBN: 0-87353-260-0]; *Level 3,* ii + 126 pp, $9 (P). [ISBN: 0-87353-261-9] Contains research projects for middle school (*Level 2*) and secondary school (*Level 3*). Units in each book give several possible projects on a topic along with a list of references. JNC

**Education, P.** *Preparing Elementary School Mathematics Teachers: Readings from the Arithmetic Teacher.* Ed: Joan Worth. NCTM, 1987, vi + 169 pp, $10 (P). [ISBN: 0-87353-251-1] Forty-eight articles, most published in the last decade, on pre-service programs, content and methods courses, and attitudes of pre-service teachers. Research summaries, reading lists, ideas for teaching specific topics, and materials and techniques for teaching. Missing discussions of mathematics curriculum, educational change, cognitive psychology, and evaluation. MW

**Education, P.** *Solving Word Problems in the Primary Grades: Addition and Subtraction.* Miriam M. Feinberg. NCTM, 1988, iii + 35 pp, $7 (P). [ISBN: 0-87353-255-4] A sequence of lessons each of which focuses on a specific concept; includes a reproducible practice page in each lesson. JNC

**Education, T(17-18), S.** *Teaching Mathematics in Grades K-8: Research Based Methods.* Ed: Thomas R. Post. Allyn & Bacon, 1988, xx + 467 pp, $37. [ISBN: 0-205-11076-2] Active researchers author chapters on various topics of elementary school mathematics including problem solving, estimation, visual thinking, and proportional reasoning as well as standard computational topics. Cognitive philosophical orientation. Not a first methods text, but ideal introduction to research for experienced teachers and beginning graduate students. Useful for secondary teachers as well. MW

**Education, S, L.** *The Ideas of Algebra, K-12, 1988 Yearbook.* Arthur F. Coxford, Albert P. Shulte. NCTM, 1988, viii + 248 pp, $16. [ISBN: 0-87353-250-3] Thirty-four articles discussing issues and strategies in teaching algebra meaningfully, including development of algebraic readiness in elementary school. Variables and function concept are central themes, and teaching with computers and calculators receives much attention. Several short articles give tips for teaching specific topics. MW

**History, L.** *The Emergence of Number.* J.N. Crossley. World Scientific, 1987, x + 222 pp, $39; $23 (P). [ISBN: 9971-50-413-8; 9971-50-414-6] Traces the origins and early development of the natural numbers, irrational numbers, and complex numbers. Many quotes from the original authors of mathematical works are included. An interesting introduction to

the history of the development of the concept of number. RH

**History, P, L**.** *Chinese Mathematics: A Concise History.* LĭYăn, Dù Shírán. Transl: John N. Crossley, Anthony W.-C. Lun. Clarendon Pr, 1987, xiii + 290 pp, $49.95. [ISBN: 0-19-858181-5] The first English translation of an important Chinese view of the history of mathematics in China, beginning with the legend of LìShŏu in 2600 B.C., continuing through the first Chinese translation of Euclid in the Míng Dynasty (1600 A.D.), and concluding with the adoption near the end of the Qīng Dynasty of Chinese translations of nineteenth-century American texts on algebra, geometry, trigonometry, and analytic geometry. Unlike the geometric inclination of classical Greek mathematics, indigenous ancient Chinese mathematics was strongly algebraic. LAS

**History.** *Capitalism and Arithmetic.* Frank J. Swetz. Open Court, 1987, xviii + 345 pp, $16.95 (P). [ISBN: 0-8126-9014-1] A socio-economic analysis of the impact that arithmetic algorithms (especially long multiplication and long division) had on commerce in 15th century Venice. Centerpiece, occupying nearly half the volume, is the first English translation (by David Eugene Smith) of the earliest printed book on arithmetic, *Treviso Arithmetic,* a practical handbook written in the Venetian dialect to help communicate the essential mathematics of commerce to a broad audience. LAS

**History, P, L.** *H.A. Kramers: Between Tradition and Revolution.* M. Dresden. Springer-Verlag, 1987, xxiv + 563 pp, $73. [ISBN: 0-387-96282-4] A biography of the Dutch theoretical physicist H.A. Kramers (1894-1952). Explores his personal relationships with other great physicists of his day (e.g., Einstein, Bohr) as well as his scientific work. AO

**Graph Theory, T(17: 2), P, L.** *Planar Graphs: Theory and Algorithms.* T. Nishizeki, N. Chiba. Annals of Disc. Math., V. 32. North-Holland (US Distr: Elsevier Science), 1988, xiii + 232 pp, $100. [ISBN: 0-444-70212-1] Introduction to planar graphs; emphasis is on constructive proofs of theorems, from which algorithms then follow. May be used as text for a course in algorithms, graph theory, or planar graphs. No exercises; algorithms given in "Pidgin" Pascal; extensive references. LC

**Combinatorics, T(16-17: 2), P, L.** *Combinatorial Designs.* W.D. Wallis. Pure & Appl. Math., V. 118. Marcel Dekker, 1988, vii + 329 pp, $89.75. [ISBN: 0-8247-7942-8] Introduction to the theory of combinatorial designs, covering the standard topics such as block designs, balanced and $t$-designs, finite geometries and Latin squares, plus some modern extensions such as one-factorizations and Room squares. Includes a chapter on finite fields. Includes exercises, solutions, hints. LC

**Combinatorics, T(16-17: 2), P, L.** *Combinatorial Configurations: Designs, Codes, Graphs.* Vladimir D. Tonchev. Transl: Robert A. Melter. Pitman Mono. & Surv. in Pure & Appl. Math., V. 40. Longman Scientific & Technical (US Distr: Wiley), 1988, 189 pp, $77.95. [ISBN: 0-582-99483-7] Chapter one presents definitions and properties of designs and symmetric block designs; also includes facts from finite fields, finite geometries and permutation groups and their connection to designs. Chapter two discusses the relationship between codes and designs, touching upon, for example, Hadamard matrices, constant weight codes and the Assmus-Mattson theorem. Finally, the last chapter covers strongly regular graphs. Exercises are interspersed throughout the text. LC

**Combinatorics, T(17-18: 1), S, P, L.** *Basic Hypergeometric Series and Applications.* Nathan J. Fine. Math. Surveys & Mono., No. 27. AMS, 1988, xiii + 124 pp, $39. [ISBN: 0-8218-1524-5] Combinatorial applications of hypergeometric series to partitions, Ramanujan's mock theta-functions, sums of squares, and modular equations. Chapter notes and references by George Andrews. BC

**Combinatorics, T(16-17: 2), L.** *Design Theory.* D.R. Hughes, F.C. Piper. Cambridge U Pr, 1988, viii + 240 pp, $17.95 (P). [ISBN: 0-521-35872-8] Corrected paperback edition of 1985 text (TR, May 1986). LC

**Number Theory, T(13-16: 1), S, L*.** *Number Theory and Its History.* Oystein Ore. Dover, 1988, x + 370 pp, $8.95 (P). [ISBN: 0-486-65620-9] Reprinting of an elementary classic, first published in 1948, based on a popular Yale course for students with minimum mathematical background. Introduces a selection of important results of number theory in the context of a historical exposition from ancient texts through Fermat's, Wilson's, and Euler's theorems to the nineteenth-century solutions of the three classic construction problems. Few problems—barely enough for class use. LAS

**Number Theory, P.** *Introduction to Analytic Number Theory.* A.G. Postnikov. Transl. of Math. Mono., V. 68. AMS, 1988, vi + 320 pp, $114. [ISBN: 0-8218-4521-7] Detailed exposition of theorems, both new and used, in additive number theory. Includes some probabilistic treatment. Note the price. BC

**Number Theory, P.** *Beilinson's Conjectures on Special Values of L-Functions.* Ed: M. Rapoport, N. Schappacher, P. Schneider. Perspectives in Math., V. 4. Academic Pr, 1988, xxiii + 373 pp, $37.50. [ISBN: 0-12-581120-9] The value of $s = 1$ of the $L$-function of a number field (Dedekind zeta function) reflects the class number and unit group of its ring of integers. The Birch and Swinnerton-Dyer conjectures assert the value at $s = 1$ of the $L$-function of a rational elliptic curve reflects the group of rational points on the curve. Beilinson's conjectures include the two as special cases. The book is a supplemental report of talks given at Oberwolfach in April 1986. GG

**Number Theory, P.** *Analytic Properties of Automorphic L-Functions.* Stephen Gelbart, Freydoon Shahidi. Perspectives in Math., V. 6. Academic Pr, 1988, vii + 131 pp, $18.75. [ISBN: 0-12-279175-4] Survey of recent work with attention focused on the general conjecture of Langlands. SB

**Number Theory, P.** *Lecture Notes in Mathematics-1301: Periods of Hecke Characters.* Norbert Schappacher. Springer-Verlag, 1988, xv + 160 pp, $17.30 (P). [ISBN: 0-387-18915-7] An exposition of recent works of Deligne, Anderson, and others on the theory of motives and its connection with periods of Hecke characters. SG

**Number Theory, P.** *Galois Representations and Arithmetic Algebraic Geometry.* Ed: Y. Ihara. Adv. Stud. in Pure Math. V. 12. North-Holland (US Distr: Elsevier Science), 1987, 373 pp, $168.50. [ISBN: 0-444-70315-2] Nineteen, mostly original, papers from a two-part symposium held October 1985 at Kyoto University, and January 1986 at the University of Tokyo. The papers are divided into four categories: Gauss Sums, Jacobi Sums, Circular Units; Braid Groups and Galois Representations; Curves and Abelian Varieties; and Class Field Theory, General Ramification Theory. GG

**Linear Algebra, T(14: 1).** *An Introduction to Linear Algebra With Applications, Second Edition.* Steven Roman. Harcourt Brace Jovanovich, 1988, viii + 504 pp, $29. [ISBN: 0-15-542736-9] The first three chapters on matrix computations in the *First Edition* (TR, December 1985) have been condensed in this new edition. Several new applications, examples, and exercises have been added throughout the text. GG

**Linear Algebra, T(16-17: 1), L.** *Nonnegative Matrices.* Henryk Minc. Wiley, 1988, xiii + 206 pp, $39.95. [ISBN: 0-471-83966-3] The core of this book is an up-to-date exposition of classical Perron-Frobenius theory together with some more recent related results. It also includes chapters on the combinatorial properties of non-negative matrices, the theory of doubly stochastic matrices, and special classes of non-negative matrices (e.g., stochastic matrices, $M$-matrices). Presumes only a background in elementary linear algebra. AO

**Linear Algebra, S*(14).** *Guide to Linear Algebra.* David A. Towers. Macmillan Education, 1988, x + 210 pp, £8.95 (P). [ISBN: 0-333-43627-X] A compact and lucid introduction to linear algebra which could serve as an excellent second source for students; includes examples and exercises. JNC

**Linear Algebra, P.** *Seven Papers Translated from the Russian.* I.V. Kovalishina, V.P. Potapov. Transl. Ser. 2, V. 138. AMS, 1988, vii + 77 pp, $40. [ISBN: 0-8218-3114-3] A series of studies of analytic matrix-valued functions in seven related papers. LAS

**Linear Algebra, T(14: 1).** *A Course in Linear Algebra.* David B. Damiano, John B. Little. Harcourt Brace Jovanovich, 1988, xiii + 434 pp, $30. [ISBN: 0-15-515134-7] A theoretic approach which defines vector spaces and linear transformation at the outset and introduces computational techniques as needed. Includes mathematical applications such as linear differential equations, as well as solutions to the majority of the exercises. JNC

**Linear Algebra, T*(15: 1, 2), S, L.** *A Primer on Linear Algebra.* I.N. Herstein, David J. Winter. Macmillan, 1988, xx + 563 pp [ISBN: 0-02-353953-4]; *Matrix Theory and Linear Algebra.* I.N. Herstein, David J. Winter. Macmillan, 1988, xviii + 508 pp. [ISBN: 0-02-353951-8] *A Primer* is a comprehensive linear algebra text which has a lot of really good problems. It is well written but yet more sophisticated than the average sophomore-level text. Its companion volume *Matrix Theory* is similar but more comprehensive, including several more advanced topics such as Jordan canonical forms. CEC

**Group Theory, T(17-18: 1), S.** *Lie Groups, Lie Algebras, and Cohomology.* Anthony W. Knapp. Math. Notes, V. 34. Princeton U Pr, 1988, xii + 509 pp, $29.50 (P). [ISBN: 0-691-08498-x] Algebraic treatment of representation theory of Lie groups, in particular of unitary groups. Expands upon one-semester graduate course whose goal was to develop material leading to the Borel-Weil-Bott theorem. Prerequisites are advanced linear algebra, metric spaces, and integration on topological groups. Includes ten pages of historical notes. GG

**Group Theory, P.** *Geometry of Group Representations.* Ed: William M. Goldman, Andy R. Magid. Contemp. Math., V. 74. AMS, 1988, xv + 312 pp, $30 (P). [ISBN: 0-8218-5082-2] Proceedings of AMS-IMS-SIAM Joint Summer Research Conference at Boulder, Colorado, July 5-11, 1987 concerning spaces of representations. SB

**Group Theory, P.** *Lattice-Ordered Groups, An Introduction.* Marlow Anderson, Todd Feil. D Reidel (US Distr: Kluwer), 1988, ix + 190 pp, $44. [ISBN: 90-277-2643-4] A lattice-ordered group or $l$-group is a group with a partial order which is in fact a lattice. This book covers the classical theory of lattice ordered groups. Emphasis is algebraic rather than analytic. Assumes only graduate courses in analysis, algebra, and point-set topology. RH

**Group Theory, P.** *Geometries and Groups.* Ed: M. Aschbacher, A.M. Cohen, W.M. Kantor. D Reidel (US Distr: Kluwer), 1988, vii + 542 pp, $118. [ISBN: 90-277-2623-X] A collection of 21 papers presented at a workshop on Geometries and Groups held in the Netherlands in 1986. Organized into four areas of interest: diagram geometries and chamber systems; incidence systems; Chevalley groups; and graphs and groups. JS

**Algebra, T(15: 1), S, L.** *Elements of Modern Algebra, Second Edition.* Jimmie Gilbert, Linda Gilbert. Ser. in Math. Prindle, Weber & Schmidt, 1988, xii + 356 pp, $30.50. [ISBN: 0-534-91502-7] Major changes in this edition include answers to half the computational exercises, more emphasis on composition of mappings, expanded treatment of induction, a section on maximal ideals, and solutions of cubic and quartic polynomials. A readable text with lots of reasonable exercises, but not as deep as other beginning algebra texts. (*First Edition,* TR, August-September 1984.) CEC

**Algebra, S(18), P.** *Noncommutative Noetherian Rings.* J.C. McConnell, J.C. Robson. Wiley, 1988, xv + 596 pp, $138. [ISBN: 0-471-91550-5] After a

brief review, the first part of the book treats general material, including Goldie's theorem. The last three parts, titled Dimensions, Extensions, and Examples, treat these aspects in more depth. Suitable for an algebraist who has taken a one-year graduate course. GG

**Algebra, P.** *Lattices with Unique Complements.* V.N. Saliĭ. Transl. of Math. Mono., V. 69. AMS, 1987, ix + 113 pp, $51. [ISBN: 0-8218-4522-5] The main problem addressed is: In the class of complete lattices, do there exist non-distributive uniquely complemented lattices? The first chapter introduces the basic results from the theory of lattices and ordered sets needed to study the main problem. The book ends with ten problems. LC

**Algebra, T(18: 1), P.** *Additive Groups of Rings, V. II.* S. Feigelstock. Pitman Res. Notes in Math. Ser., V. 169. Longman Scientific & Technical (US Distr: Wiley), 1988, 100 pp, $39.95 (P). [ISBN: 0-582-01370-4] The second volume presents strictly ring-theoretic results proven using the additive structures of the rings, and further theory of additive groups of rings (continuing *Volume I*). Embedding theorems, radical theory, fissible rings, subring-ideal-quotient ring properties, subgroups which are ideals in every ring, miscellaneous results. No exercises. RB

**Algebra, P.** *Lecture Notes in Mathematics-1300: Constructions of Lie Algebras and their Modules.* George B. Seligman. Springer-Verlag, 1988, vi + 190 pp, $17.30 (P). [ISBN: 0-387-18973-4] Studies constructing irreducible $g$-modules of finite dimension, where $g$ is a central simple Lie algebra over a field $F$ of characteristic zero. In particular, demonstrates constructions of the irreducible modules in the isotropic cases with non-reduced root systems, and in all anisotropic cases where $[g : F]$ is not among 14, 28, 52, 78, 133, or 248. RH

**Algebra, P.** *Lecture Notes in Mathematics-1328: Ring Theory.* Ed: J.L. Bueso, P. Jara, B. Torrecillas. Springer-Verlag, 1988, ix + 331 pp, $28.60 (P). [ISBN: 0-387-19474-6] Proceedings of a conference held in Granada, Spain, September 1-6, 1986. SB

**Algebra, T(17), L.** *Algebra, Third Edition.* Saunders MacLane, Garrett Birkhoff. Chelsea, 1988, xix + 626 pp, $28.50. [ISBN: 0-8284-0330-9] A classic in the axiomatic approach in algebra. Only major change from *Second Edition* is the inclusion of a chapter on affine and projective geometry. Excellent reference; covers almost everything. (1967 Macmillan *First Edition,* TR, November 1967; Extended Review, January 1971; 1979 Macmillan *Second Edition,* TR, October 1979.) MR

**Algebra, T**(15: 1, 2), S*, L*.** *A First Undergraduate Course in Abstract Algebra, Fourth Edition.* Abraham P. Hillman, Gerald L. Alexanderson. Wadsworth, 1988, xiv + 541 pp. [ISBN: 0-534-08844-9] The sections on Boolean algebras and their applications has been thoroughly revised in this edition. The treatment of Euclid's algorithm has been rewritten in light of the new emphasis on algorithms, and material on dihedral groups has been added. Minor changes appear throughout the book. A readable and useable text with a wealth of exercises. (*First Edition,* TR, August-September 1974; *Second Edition,* TR, June-July 1978; *Third Edition,* TR, March 1983.) CEC

**Algebra, T*(17: 1, 2), S*, P*, L*.** *Ring Theory, Volume I.* Louis H. Rowen. Pure & Appl. Math., V. 127. Academic Pr, 1988, xxiv + 538 pp, $89.50. [ISBN: 0-12-599841-4] A sourcebook of the results and proofs which make up the general structural theory of rings. Chapter titles include the construction of rings, basic structural theory, rings of fractions, embedding theorems, and categorical aspects of module theory. A wide-ranging synthesis which is a welcome addition to the literature. Written in a formal style. Includes an excellent set of exercises and a comprehensive bibliography. CEC

**Algebra, T(14-15: 1), S*, L.** *Guide to Abstract Algebra.* Carol Whitehead. Math. Guides. Macmillan Education, 1988, xiii + 257 pp, £8.95 (P). [ISBN: 0-333-42657-6] Covers sets, relations, mappings, integers, basic group theory. Many examples and exercises with solutions. SB

**Algebra, T(13: 1).** *Algebra for College Students, Second Edition.* M.A. Munem, W. Tschirhart. Worth, 1988, xii + 609 pp, $28.95. [ISBN: 0-87901-384-2] A straightforward presentation of algebra topics (including functions, systems of equations and inequalities, exponential and logarithmic functions, and topics in discrete mathematics); this edition notes the algebraic changes which produce shifting, reflecting, and stretching of graphs. JNC

**Calculus, T*(13-14: 3).** *Calculus with Analytic Geometry.* Richard A. Hunt. Harper & Row, 1988, xx + 1176 pp. [ISBN: 0-06-043036-2] No historical remarks; no biographical sketches; no footnotes; no appendices; no tables; no asterisked problems requiring a calculator; no skin-deep applications in economics, biology, or psychology. What there is is readable, informal, clear but not wordy. Emphasises understanding of concepts over memorization of formulas. About 1.5 computer-generated figures per page. Parabolas and hyperbolas look like they should. Excellent pictures of surfaces. Answers have been checked and re-checked. For science and engineering students in a three-semester calculus sequence. Available are an instructor's solution manual, a student solutions manual; free to instructors: transparencies, a computerised test-generation system, and a set of six diskettes for tutorial purposes. JK

**Calculus, T(13-14: 1).** *Calculus for the Management, Life, and Social Sciences, Second Edition.* Bernard Kolman, Charles G. Denlinger. Harcourt Brace Jovanovich, 1988, xi + 676 pp, $30. [ISBN: 0-15-505754-5] Major changes in this edition include moving some of the early review material to an appendix, and addition of material on functions, more business applications, additional exercises, new sections on the Newton-Raphson method, on numerical integration, and on differentials and their applica-

tions. (*First Edition*, TR, November 1981.) RH

**Calculus, T(13: 2).** *Calculus and Analytic Geometry, Brief, Fourth Edition.* Sherman K. Stein. McGraw-Hill, 1987, xviii + 748 pp, $38.95. [ISBN: 0-07-061162-9] Consists of the first ten chapters of *Calculus and Analytic Geometry, Fourth Edition.* (TR, *First Edition*, November 1973; Extended Review, February 1976; *Third Edition*, TR, October 1982; *Fourth Edition*, TR, April 1987.) JNC

**Calculus, T*(14-15: 1, 2), L.** *Vector Calculus, Third Edition.* Jerrold E. Marsden, Anthony J. Tromba. WH Freeman, 1988, xiv + 655 pp, $38.95. [ISBN: 0-7167-1856-1] Major changes in this latest edition (*First Edition*, TR, March 1976; Extended Review, May 1977; *Second Edition*, TR, November 1981) is a much larger and more diverse set of exercises backed up by a separate *Study Guide* with complete solutions. Improved exposition, some reorganization, and many illustrations make the text more useful for a wide variety of students. LAS

**Calculus, T(13: 2).** *Calculus, Second Edition.* Dennis D. Berkey. Saunders College, 1988, xviii + 1122 pp, $41. [ISBN: 0-03-008899-2] Changes in this edition include extensive rewriting of many sections, some reorganisation, shortening of the precalculus review; addition of a chapter on antidifferentiation before coverage of the definite integral, consolidation into one chapter of the differential equation topics, and a new four-color insert depicting three-dimensional surfaces. (*First Edition*, TR, December 1984.) JNC

**Calculus, T(13-14: 1, 2).** *Calculus with Applications.* Karl J. Smith. Brooks/Cole, 1988, xiii + 396 pp, $30. [ISBN: 0-534-08898-8] An updated text for business, management, life, or social sciences students. "Emphasis throughout is to enhance students' understanding of the modeling process and how mathematics is used in real world applications." Integration techniques (except integration by parts, tables) omitted in favor of numerical methods, more applications (e.g., probability), calculators, muMath. Computerized test bank. RB

**Calculus, T*(16-17: 1), P, L.** *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Jan R. Magnus, Heinz Neudecker. Prob. & Math. Stat. Wiley, 1988, xvii + 393 pp, $51.95. [ISBN: 0-471-91516-5] Self-contained and unified treatment of matrix differential calculus for students and practitioners of econometrics and statistics. Approach is via differentials which authors believe to be "more congenial to multivariable functions as they crop up in econometrics, mathematical statistics or psychometrics than derivatives." Exercises complement text. Bibliography and bibliographical notes. Applications. JK

**Real Analysis, P.** *Fractional Calculus: Integrations and Differentiations of Arbitrary Order.* Katsuyuki Nishimoto. Descartes Pr. *Volume I*, 1984, xiii + 195 pp; *Volume II*, 1987, xiv + 189 pp. A two-volume monograph on theory and applications of fractional-order differentiation and integration—"differintegration," in the author's terminology—of functions of one and several variables. Rich in examples and graphs. The use of English is slightly non-standard. PZ

**Complex Analysis, S(18), P.** *Lecture Notes in Mathematics-1307: A Real Variable Method for the Cauchy Transform, and Analytic Capacity.* Takafumi Murai. Springer-Verlag, 1988, vii + 133 pp, $13.90 (P). [ISBN: 0-387-19091-0] A study of analytic capacity, especially of discontinuous curves, graphs, and "cranks." Properties of analytic capacity are studied via connections with the real-variable theory of the Cauchy-Hilbert transform, and with integral-geometric properties of the curves in question. PZ

**Complex Analysis, P.** *Complex Analysis, Functional Analysis and Approximation Theory.* Ed: Jorge Mujica. Math. Stud., V. 125. North-Holland (US Distr: Elsevier Science), 1986, viii + 297 pp, $44.50 (P). [ISBN: 0-444-87997-8] Proceedings of a conference held at Universidade Estadual de Campinas, Brazil, July 1984. MR

**Complex Analysis, S(18), P.** *The Complex Analytic Theory of Teichmüller Spaces.* Subhashis Nag. Wiley, 1988, xiii + 427 pp, $54.95. [ISBN: 0-471-62773-9] An introduction to the fundamental facts concerning the complex analytic structure of Teichmüller spaces and other moduli spaces which should be accessible to graduate students. Central fact presented is that the Teichmüller space, parametrizing complex structures on a given surface, itself carries the complex structure of a complex manifold. RH

**Complex Analysis, S(18), P.** *Methods of Complex Analysis in Partial Differential Equations with Applications.* Manfred Kracht, Erwin Kreyszig. Canadian Math. Soc. Ser. of Mono. & Adv. Texts. Wiley, 1988, xiv + 394 pp, $54.95. [ISBN: 0-471-83091-7] An inviting, readable study of complex analysis methods—especially complex integral operator methods—applied to problems in partial differential equations. Detailed, concrete applications of the theory are made to three classes of partial differential equations of interest to mathematical physicists. Includes an enormous bibliography. PZ

**Complex Analysis, P.** *Lecture Notes in Mathematics-1312: Analytic Functions Smooth up to the Boundary.* Nikolai A. Shirokov. Springer-Verlag, 1988, 213 pp, $21.20 (P). [ISBN: 0-387-19255-7] A study of the theory of Nevanlinna, or inner-outer factorization of functions of one complex variable, analytic in the unit disc and smooth up to the boundary. Includes a useful historical introduction. PZ

**Differential Equations, P.** *Multiphase Averaging for Classical Systems: With Applications to Adiabatic Theorems.* P. Lochak, C. Meunier. Appl. Math. Sci., V. 72. Springer-Verlag, 1988, xi + 360 pp, $39.80 (P). [ISBN: 0-387-96778-8] Averaging out "fast" variables is sometimes necessary to see what the slowpokes are doing. Complete statements and proofs of hard-to-find (mostly Russian) results are presented. BC

**Differential Equations, T(14: 1).** *Differential Equations: A First Course, Second Edition.* Martin M. Guterman, Zbigneiw H. Nitecki. Saunders College, 1988, x + 742 pp, $34. [ISBN: 0-03-0096170] Standard topics for engineering and science students: first-order equations, linear equations, linear systems, Laplace transform, numerical approximation, partial differential equations, and Fourier series. Consideration of physical models provides motivation. This edition includes rewritten material, new material (e.g., on phase portraits, linearization near equilibria, stability), additional exercises (many of an applied nature). (*First Edition,* TR, November 1984.) DFA

**Differential Equations, P.** *Compactness Methods for Nonlinear Evolutions.* I.I. Vrabie. Pitman Mono. & Surv. in Pure & Appl. Math., V. 32. Longman Scientific & Technical (US Distr: Wiley), 1987, 325 pp, $77.95. [ISBN: 0-470-20738-8] Recent results in existence theory for nonlinear evolution equations with "non-accretive" operators. Applications to partial differential equations. BC

**Differential Equations, T***(16-17), L.** *Geometrical Methods in the Theory of Ordinary Differential Equations, Second Edition.* V.I. Arnold. Transl: Joseph Szücs. Grund. der math. Wissenschaften, B. 250. Springer-Verlag, 1988, xiii + 351 pp, $48. [ISBN: 0-387-96649-8] A beautiful book by one of the founders of the theory of geometrical analysis of differential equations. Accessible to all mathematicians and physicists. In some ways a sequel to his classic *Ordinary Differential Equations.* Style of presentation here is similar: the focus is on simple examples explored in detail to reveal deep and complex ideas. Full of Arnold's unique ingenious problems and examples. Covers stability, perturbation theory, normal forms, and local bifurcation theory. (*First Edition,* TR, January 1984.) MR

**Differential Equations, P.** *Mathematics Applied to Science: In Memoriam Edward D. Conway.* Ed: Jerome Goldstein, Steven Rosencrans, Gary Sod. Academic Pr, 1988, xx + 309 pp, $34.50. [ISBN: 0-12-289510-X] Papers from a November 1986 conference honoring Edward Conway, who died suddenly the previous year. 13 papers on diverse applications of differential equations, together with biographical information on Conway who had been influential in leading Tulane into the computer age. LAS

**Partial Differential Equations, P.** *Recent Topics in Nonlinear PDE III.* Ed: Kyûya Masuda, Takashi Suzuki. Math. Stud., V. 148. North-Holland (US Distr: Elsevier Science), 1987, vii + 266 pp, $110.50 (P). [ISBN: 0-444-70317-9] An outgrowth of lectures delivered at the fourth meeting on "Nonlinear Partial Differential Equations," held at the University of Tokyo, February 24-26, 1986. There are ten chapters discussing nonlinear partial differential equations occurring in the area of fluid dynamics, free boundary problems, population dynamics, and mathematical physics. RSF

**Partial Differential Equations, P.** *Nonlinear Partial Differential Equations and Their Applications: Collège de France Seminar, Volume VIII.* Ed: H. Bresis, J.L. Lions. Longman Scientific & Technical (US Distr: Wiley), 1988, 220 pp, $51.95 (P). [ISBN: 0-470-20993-3] Contains the texts of selected lectures delivered by leading international experts at a weekly seminar held at Collège de France. The main theme is recent work in nonlinear partial differential equations. The emphasis is on applications including control theory, theoretical physics, fluid and continuum mechanics, free boundary problems, dynamical systems, scientific computing, numerical analysis and engineering. In English and French. CEC

**Partial Differential Equations, P.** *Energy Stability and Convection.* Ed: G.P. Galdi, B. Straughan. Pitman Res. Notes in Math. Ser., V. 168. Longman Scientific & Technical (US Distr: Wiley), 1988, 448 pp, $74.95 (P). [ISBN: 0-582-00318-0] Nine papers and eight abstracts, from a May 1986 meeting on the Island of Capri. MR

**Partial Differential Equations, P.** *Huygens' Principle and Hyperbolic Equations.* Paul Günther. Perspectives in Math., V. 5. Academic Pr, 1988, lvii + 847 pp, $69. [ISBN: 0-12-307330-8] Huygens' principle is a property of hyperbolic operators in the equations for wave propagation in curved space-time (first fully formulated by Hadamard, not Huygens). Emphasis is on finding, describing, and classifying examples of nontrivial Huygens operators. BC

**Partial Differential Equations, P.** *Multigrid Solution of the Steady Euler Equations.* S.P. Spekreijse. CWI Tract, V. 46. Math Centrum, 1988, 153 pp, Dfl. 24.20 (P). [ISBN: 90-6196-346-X] Uses discretisations based on cell-centered finite volume schemes. Develops robust and efficient methods to solve these. DFA

**Numerical Analysis, P.** *Regularization for Applied Inverse and Ill-Posed Problems: A Numerical Approach.* Bernd Hofmann. Teubner-Texte zur Math., B. 85. BG Teubner Leipzig, 1986, 196 pp, 19 M (P). [ISBN: 3-322-00341-8] Discusses numerical methods for the solution of linear and nonlinear inverse and ill-posed problems of the type that arise in scientific and engineering applications. All the methods presented are based on regularization of a discretized inverse problem. AO

**Numerical Analysis, T(17-18: 2), L.** *Theory of Difference Equations: Numerical Methods and Applications.* V. Lakshmikantham, D. Trigiante. Math. in Sci. & Engin., V. 181. Academic Pr, 1988, x + 242 pp, $49.95. [ISBN: 0-12-434100-4] Introduction to difference equations, with attention toward iterative processes and numerical methods for differential equations. Not covered in detail: classical applications to interpolation, numerical quadrature, and differentiation. Includes a chapter on examples of real world models such as population dynamics and the cobweb model (from economics). LC

**Numerical Analysis, T(16-17: 1), S, P, L.** *Computational Methods for Integral Equations.* L.M.

Delves, J.L. Mohamed. Cambridge U Pr, 1988, xii + 376 pp, $24.95 (P). [ISBN: 0-521-35796-9] Considers practical solutions to one-dimensional integral equations, including Fredholm and Volterra equations of the first, second, and third kind; some background in numerical methods and linear algebra is assumed. There are 14 chapters, an appendix on singular expansions, a comprehensive reference list, chapter exercises, and numerous examples; numerical results are presented (but there are no program codes). Attention is paid to both theory of solutions and rates of convergence achievable for numerical approximations. (1985 *First Edition*, TR, December 1986.) RSF

**Numerical Analysis, P.** *Numerical Algorithms for Modern Parallel Computer Architectures.* Ed: Martin Schultz. IMA, V. 13. Springer-Verlag, 1988, xi + 232 pp, $25. [ISBN: 0-387-96733-8] Papers from a workshop on Scientific Computation at the IMA in Minneapolis, Minnesota, 1986-87, concerned with multiprocessor architectures, parallel systems, algorithms, and languages, and the application of massively parallel architectures (and the new approaches to applied mathematical modelling and scientific computation they will stimulate) to real world problems. RM

**Numerical Analysis, T(18), P.** *Numerical Grid Generation: Foundations and Applications.* Joe F. Thompson, Z.U.A. Warsi, C. Wayne Mastin. North-Holland (US Distr: Elsevier Science), 1985, xv + 483 pp, $34.95 (P). [ISBN: 0-444-00985-X] Numerical grid generation is commonly used in the solution of partial differential equations and in finite element methods over arbitrary regions. Seeks to provide the mathematical background and application techniques necessary to use numerical grid generation effectively in numerical software. SM

**Operator Theory, P.** *Index Theory of Elliptic Operators, Foliations, and Operator Algebras.* Ed: Jerome Kaminker, Kenneth C. Millett, Claude Schochet. Contemp. Math., V. 70. AMS, 1988, x + 322 pp, $32 (P). [ISBN: 0-8218-5077-6] Proceedings of AMS Special Sessions held January 7, 1986 in New Orleans, and April 11, 1986 in Indianapolis. LC

**Operator Theory, T(18: 2).** *Interpolation of Operators.* Colin Bennett, Robert Sharpley. Pure & Appl. Math., V. 129. Academic Pr, 1988, xiv + 469 pp, $69.95. [ISBN: 0-12-088730-4] Develops theory of the real method of interpolation through the theory of spaces of measurable functions. Assumes familiarity with the fundamentals of real analysis, measure theory, and functional analysis. LC

**Functional Analysis, P.** *Amenable Banach Algebras.* Jean-Paul Pier. Pitman Res. Notes in Math. Ser., V. 172. Longman Scientific & Technical (US Distr: Wiley), 1988, 161 pp, $47.95 (P). [ISBN: 0-470-21066-4] Amenability, first studied on discrete groups, is studied here on Banach algebras, and more specifically for $C^*$-algebras and von Neumann algebras. AWR

**Functional Analysis, T(18: 2), P.** *Spectral Theory and Differential Operators.* D.E. Edmunds, W.D. Evans. Math. Mono. Clarendon Pr, 1987, xvii + 574 pp, $115. [ISBN: 0-19-853542-2] A study of the relationship between the classical theory of compact operators in Banach or Hilbert spaces and boundary-value problems for elliptic differential equations. Considers the results of the last decade. Includes a substantial list of references. CEC

**Functional Analysis, S(18), P.** *Representations of *-Algebras, Locally Compact Groups, and Banach *-Algebraic Bundles, Volume 1 and 2.* J.M.G. Fell, R.S. Doran. Academic Pr, 1988, $99 each. *Volume 1: Basic Representation Theory of Groups and Algebras.* Pure & Appl. Math., V. 125. xviii + 746 pp [ISBN: 0-12-252721-6]; *Volume 2: Banach *-Algebraic Bundles, Induced Representations, and the Generalized Mackey Analysis.* Pure & Appl. Math., V. 126. viii + 739 pp. [ISBN: 0-12-252722-4] First volume and part of second volume could be used as a "leisurely introduction to the basic functional analysis underlying the theory of infinite-dimensional unitary representations of locally compact groups and its generalisation to representations of Banach algebras." Material selected to lead into Mackey normal subgroup analysis in context of Banach *-algebraic bundles, which comprise remainder of second volume. First author's own research is included here. Each chapter concludes with exercises and notes. GG

**Analysis, P, L.** *Rational Approximation of Real Functions.* P.P. Petrushev, V.A. Popov. Encylop. of Math. & Its Applic., V. 28. Cambridge U Pr, 1987, xi + 371 pp, $69.50. [ISBN: 0-521-33107-2] A survey of the basic results in the field including recent work. Some results from the theory of linear approximation and from the theory of complex rational approximation are included for the sake of completeness and comparison. AO

**Analysis, P.** *Nonlinear Analysis.* Ed: Th. M. Rassias. World Scientific, 1987, xi + 557 pp, $55. [ISBN: 9971-50-140-6] Twenty papers on applications of topological methods, fixed-point and bifurcation theory, and stability of mapping to nonlinear analysis. Includes a lengthy survey of theoretical and experimental developments in bifurcation and nonlinear instability problems. BC

**Analysis, T(18), P.** *Approximation of Continuously Differentiable Functions.* José G. Llavona. Math. Stud., V. 130. North-Holland (US Distr: Elsevier Science), 1986, xiv + 241 pp, $48 (P). [ISBN: 0-444-70128-1] Presents results, classical and modern, about approximation of continuously differentiable functions on real manifolds locally of finite dimension and on Banach spaces. Much of the material has not appeared in book form previously. LC

**Analysis, P.** *Lecture Notes in Mathematics-1305: Strong Asymptotics for Extremal Polynomials Associated with Weights on R.* D.S. Lubinsky, E.B. Saff. Springer-Verlag, 1988, vii + 153 pp, $17.30 (P). [ISBN: 0-387-18958-0] Following work of Freud and Nevai in the late 1960's and results of Erdős, these notes answer questions related to growth rates

of weights for Ullman distributions. AWR

**Analysis, T(18: 1, 2), S, P.** *Lecture Notes in Mathematics-1319: Conformal Geometry and Quasiregular Mappings.* Matti Vuorinen. Springer-Verlag, 1988, xix + 209 pp, $21.20 (P). [ISBN: 0-387-19342-1] A self-contained, readable introduction to quasiconformal and quasiregular mappings—generalizations to Euclidean $n$-space of conformal and analytic mappings, respectively. Based on a 1986 course of lectures at the University of Helsinki; should be accessible and useful to graduate students. Exercises are liberally sprinkled throughout the text. PZ

**Analysis, T(16-17: 1, 2).** *Nonstandard Analysis.* Alain Robert. Wiley, 1988, xx + 156 pp, $34.95. [ISBN: 0-471-91703-6] A careful introduction for advanced undergraduates and practicing teachers (who "should now learn that their practice can be reconciled with theory"). First part covers basic theory from scratch, through derivatives and integrals; second part is more advanced, covering invariant means, approximations, differential equations, Green functions, and invariant subspaces. LAS

**Algebraic Geometry, P.** *Algebraic Geometry, Sendai, 1985.* Ed: T. Oda. Adv. Studies in Pure Math., V. 10. North-Holland (US Distr: Elsevier Science), 1987, 794 pp, $236.75. [ISBN: 0-444-70313-6] Original papers by speakers and participants, twenty-six in all, from a June 1985 symposium held at Tohoku University in Sendai, Japan. The scope is wide and there is some attempt to survey other recent work. GG

**Differential Geometry, T*(17-18: 1), S, L*.** *Geometric Measure Theory: A Beginner's Guide.* Frank Morgan. Academic Pr, 1988, viii + 145 pp, $19.95. [ISBN: 0-12-506855-7] A profusely-illustrated, plain-spoken preface to standard texts (e.g., Federer), where rectifiable sets in $R^3$ are not just abstract symbols but such concrete images as the surfaces of countably many bicycles. Focus is on the search for surfaces of minimal area with given boundary. Assumes some graduate-level real analysis. Exercises, solutions, bibliography. LAS

**Differential Geometry, P.** *Differential Geometry Applied to Curve and Surface Design, Volume 1: Foundations.* Anthony W. Nutbourne, Ralph R. Martin. Halsted Pr, 1988, 282 pp, $115. [ISBN: 0-470-21036-2] Innovative approach to the synthesis of curves and surfaces. Of particular interest is the use of patches called cyclides to describe a surface. The history of cyclides can be traced to 1801; early researchers include Maxwell and Cayley. Authors claim that their approach is more precise than traditional methods when the surface to be synthesised is almost flat and must be free of imperfections (e.g., car body parts). MR

**Differential Geometry, T**(16-17: 1), L**.** *Lectures on Classical Differential Geometry, Second Edition.* Dirk J. Struik. Dover, 1988, viii + 232 pp, $6.95 (P). [ISBN: 0-486-65609-8] Unabridged and unaltered republication of the 1961 *Second Edition.* Contains an appendix on Cartan's method of Pfaffians which was not in the 1950 edition. One of my favorite books ever. Authoritative, scholarly, clear. Emphasizes geometry not vector algebra. This was the book to study while trying to understand another. Peppered with helpful asides and references to original sources. Profusely illustrated with line drawings as only Addison-Wesley provided at the time the book first appeared. Excellent exercises/problems with hints/answers to almost all. For use in a course for advanced undergraduates or beginning graduate students. Every serious undergraduate student in mathematics should own a copy of this book. JK

**Differential Geometry, T*, S, P, L*.** *Lecture Notes on Elementary Topology and Geometry.* I.M. Singer, J.A. Thorpe. Springer-Verlag, 1987, viii + 232 pp, $34. [ISBN: 0-387-90202-3] Third printing of Springer's 1977 reprinting (TR, June-July 1977) of the classic text first published by Scott Foresman in 1967 (TR, November 1967; Extended Review, February 1968). The first text to provide a unified view for good undergraduates of topology (point set theory, fundamental groups, simplicial complexes) and geometry (manifolds, DeRham theory, Riemannian geometry). LAS

**Geometry, T**, S**, L**.** *Build Your Own Polyhedra.* Peter Hilton, Jean Pedersen. Addison-Wesley, 1988, 175 pp, (P). [ISBN: 0-201-22060-1] A "must have" book for anyone who even thinks polyhedra might be interesting. Accessible to "any intelligent person aged between twelve and one hundred." Great for self-study. Super for the in-service and pre-service education of teachers. Throughout are lists of required materials, hints, and instructions with clear diagrams. The mathematical richness of the material reflects the experience and expertise of the authors. Slim, but meaty, with challenges. From folding paper strips into polygons to constructing collapsoids (pseudo-sonohedra which can be folded flat in various ways). Culminates in a chapter including puzzle-based volume calculations, Euler's formula, Descartes' (angular deficiency) formula, combinatorial properties of polyhedra, and symmetries of the cube. Spiral-bound, the book is not consumed as models are built. Don't pass on this one. JK

**Topology, T(18: 2), S, P.** *Theory of Topological Structures: An Approach to Categorical Topology.* Gerhard Preuss. Transl: Andreas Behling. Math. & Its Applic. D Reidel (US Distr: Kluwer Academic), 1988, xii + 304 pp, $69. [ISBN: 90-277-2627-2] A nearness space is a pair $(X, \mu)$ where $X$ is a set and $\mu$ is a particular sort of subcollection of covers of $X$. Every topological concept may be described in terms of nearness spaces; since nearness spaces form a topological category, one may use the techniques of category theory to study topology. Although this book is a less than fluent translation of the German text, there previously have been no books available on this relatively young subject, categorical topology. While the last chapter requires Čech cohomology theory, the first six chapters presume only some

basic facts from general topology. A collection of exercises follow the text. LW

**Operations Research, T(15-17).** *An Introduction to Linear Programming and Game Theory, Second Edition.* Paul R. Thie. Wiley, 1988, xv + 396 pp, $50.22. [ISBN: 0-471-62488-8] Emphasis is on using linear programming to solve problems. Tries to keep mathematical prerequisites to a minimum. Gives attention to mathematical modelling as an applied tool. The book is very similar to the *First Edition* (TR, June-July 1979). AWR

**Operations Research, S(18), P, L.** *Nonconvex Programming.* Ferenc Forgó. Akademiai Kiado, 1988, 188 pp, $24. [ISBN: 963-05-4453-9] Provides an overview of basic directions in current research. Emphasizes methods over theory. Extensive bibliography. Examples, no exercises. SM

**Operations Research, S(13).** *Dynamic Programming: An Elegant Problem Solver.* William Sacco, *et al.* Contemp. Appl. Math. Janson, 1987, 56 pp, $7.95 (P). [ISBN: 0-939765-05-5] Written for precalculus or freshman applied mathematics courses, this monologue presents dynamic programming through an interesting series of examples. Questions lead the reader through examples, and preview new topics. Well written. Can be used as a textbook for a short unit, or as a self-study guide. SM

**Optimisation, T(16-17: 1), S, L.** *Computer Solution of Linear Programs.* J.L. Nasareth. Mono. on Numer. Analy. Oxford U Pr, 1987, xiii + 231 pp, $39.95. [ISBN: 0-19-504278-6] This book is distinguished from most other texts on linear programming by its detailed presentation of the ideas that underlie practical implementations of the simplex algorithm. It does not presume prior knowledge of the theory of linear programming or the simplex algorithm. AO

**Optimisation, P*, L.** *Geometric Algorithms and Combinatorial Optimization.* Martin Grötschel, László Lovász, Alexander Schrijver. Algorithms & Combinatorics, V. 2. Springer-Verlag, 1988, xii + 362 pp, $68. [ISBN: 0-387-13624-X] Develops two geometric algorithms (the ellipsoid method for finding a point in a convex set, and the basis reduction method for point lattices) and uses them to prove the polynomial time solvability of several problems in combinatorial optimisation. For some of these problems, the techniques described here are the only known methods for establishing polynomial time solvability. AO

**Dynamical Systems, S(18), P, L.** *New Directions in Dynamical Systems.* Ed: T. Bedford, J. Swift. London Math. Soc. Lect. Note Ser., V. 127. Cambridge U Pr, 1988, xiii + 283 pp, $34.50 (P). [ISBN: 0-521-34880-3] Six review papers growing out of a 1986 conference held at King's College, Cambridge, intended to introduce a reader with some knowledge of dynamical systems to several frontier areas. LAS

**Dynamical Systems, T(15-17), S, P*, L**.** *The Science of Fractal Images.* Ed: Heins-Otto Peitgen, Dietmar Saupe. Springer-Verlag, 1988, xiii + 312 pp, $34. [ISBN: 0-387-96608-0] A sequel to *The Beauty of Fractals* (TR, November 1986), this monograph, based on notes from a SIGGRAPH '87 short course, focuses on the mathematical and algorithmic details required to create beautiful fractal computer graphics. Following a lengthy historical foreword by Mandelbrot, five main chapters by Voss, Saupe, Devaney, Peitgen, and Barnsley introduce a variety of algorithms for diverse fractal images, from Julia sets to landscapes. Four appendices add details for efficiency and aesthetics. A superb sourcebook for a senior seminar on fractals. LAS

**Dynamical Systems, P.** *Dynamics Reported, Volume 1.* Ed: U. Kirchgraber, H.O. Walther. Wiley, 1988, ix + 306 pp, $71.95. [ISBN: 0-471-91661-7] First in a series of volumes featuring lengthy articles that give high-level contemporary surveys of the rapidly developing frontiers of dynamical systems. This volume contains five papers on such topics as Mather sets, orbits in diffusion equations, and nonlinear resonance. LAS

**Dynamical Systems, T(16-17: 1), S, P, L.** *From Equilibrium to Chaos: Practical Bifurcation and Stability Analysis.* Rüdiger Seydel. North-Holland (US Distr: Elsevier Science), 1988, xv + 367 pp, $55. [ISBN: 0-444-01250-8] A leisurely introduction to "practical" aspects of nonlinear phenomena, emphasising numerical methods. Chapters on continuation, branching behavior, stability of periodic solutions, and chaos. BC

**Control Theory, P.** *Lecture Notes in Control and Information Sciences-106: Nonlinear Time Series and Signal Processing.* Ed: R.R. Mohler. Springer-Verlag, 1988, v + 145 pp, $21.80 (P). [ISBN: 0-387-18861-4] A sample of new results on dynamical nonlinear statistical modeling in an area where most results are confined to linear Gaussian methodologies. AWR

**Control Theory, S(18), P.** *Lecture Notes in Control and Information Sciences-107: Structural Analysis and Design of Multivariable Control Systems: An Algebraic Approach.* Y.T. Tsay, L.-S. Shieh, S. Barnett. Springer-Verlag, 1988, vi + 208 pp, $32.70 (P). [ISBN: 0-387-18916-5] Seeks to begin bridging the gap between algebraic system theory and system design methods. Analysis is based on state-space representations and the characteristic matrix polynomial of a multivariate control system. SM

**Control Theory, T(17: 1), L.** *Lecture Notes in Control and Information Sciences-88: A Course in $H_\infty$ Control Theory.* Bruce A. Francis. Springer-Verlag, 1987, x + 150 pp, $26.70 (P). [ISBN: 0-387-17069-3] An elementary introduction to the design of linear control systems to meet frequency domain performance specifications expressed in terms of an $H_\infty$-norm of a transfer matrix that represents the maximum overall frequencies of its largest singular value. LAS

**Control Theory, P.** *Lecture Notes in Control and Information Sciences-104: Frequency Domain Properties of Scalar and Multivariable Feedback Systems.* J.S. Freudenberg, D.P. Loose. Springer-Verlag, 1988, viii + 281 pp, $38.80 (P). [ISBN: 0-387-18869-X]

Study of the frequency response properties of linear time invariant multivariable feedback systems. Extension of the classical scalar theory, with emphasis on the constraints imposed on a feedback structure, by realization questions, and by properties of the controlled system. Mathematically rigorous theory, with heuristics for design. RM

**Control Theory, P.** *Lecture Notes in Control and Information Sciences-111: Analysis and Optimization of Systems.* Ed: A. Bensoussan, J.L. Lions. Springer-Verlag, 1988, xiv + 1175 pp, $132.10 (P). [ISBN: 0-387-19237-9] Contains over 100 of the papers presented to the Eighth International Conference on Analysis and Optimization of Systems. SM

**Control Theory, T(16-17: 1, 2), S.** *Elements of Finite-Dimensional Systems and Control Theory.* N.U. Ahmed. Pitman Mono. & Surv. in Pure & Appl. Math., V. 37. Longman Scientific & Technical (US Distr: Wiley), 1988, xiii + 421 pp. [ISBN: 0-470-20987-9] Introductory text requiring advanced calculus, elementary probability, and some real analysis as a background. Treats linear and nonlinear deterministic and stochastic systems. Includes stability and optimal control. Each chapter concludes with numerous exercises. GG

**Probability, P.** *Lecture Notes in Mathematics-1299: Probability Theory and Mathematical Statistics.* Ed: S. Watanabe, Yu. V. Prokhorov. Springer-Verlag, 1988, viii + 589 pp, $56.40 (P). [ISBN: 0-387-18814-2] Proceedings of the Fifth Japan-USSR Symposium on Probability Theory held in Kyoto, Japan, July 8-14, 1986. Contains 62 papers on a variety of topics. RSK

**Elementary Statistics, T(13: 1, 2).** *Statistics and Probability in Modern Life, Fourth Edition.* Joseph Newmark. Saunders College, 1988, xxi + 744 pp, $30. [ISBN: 0-03-008367-2] Revision of the author's 1983 *Third Edition* (TR, January 1984; *Second Edition,* TR, March 1978). Contents have been reorganized, updated, and expanded to make it a more attractive text. Has extensive coverage of descriptive statistics and probability, with the result that estimation and hypothesis testing are not part of the suggested syllabus for a one-semester course. RSK

**Elementary Statistics, T(15).** *Probability Modeling and Computer Simulation.* Norman S. Matloff. Prindle, Weber & Schmidt, 1988, x + 358 pp. [ISBN: 0-534-91854-9] A text for a one-term introduction to a calculus-based probability and statistics course for engineering and computer science students. Simulation methods are used in parallel with mathematical analysis to present new ideas and solve homework exercises. Students are asked to write simple Pascal programs. Many examples from engineering and computer science. Many exercises, answers to a few. SM

**Computational Statistics, P*.** *Directory of Statistical Microcomputer Software, 1988 Edition.* Wayne A. Woodward, *et al.* Marcel Dekker, 1988, iii + 744 pp, $59.75 (P). [ISBN: 0-8247-7846-4] Describes over 200 statistical programs and packages, including statistical features supported, documentation available, configurations supported, number of users, cost, etc. Does not evaluate them but gives references to reviewers when provided by the manufacturer. Appendices include tables indicating which programs perform certain statistical tasks and which run on certain systems or types of computers. RSK

**Computational Statistics, T(18), S, P*.** *Elements of Statistical Computing: Numerical Computation.* Ronald A. Thisted. Chapman & Hall, 1988, xx + 427 pp, $45. [ISBN: 0-412-01371-1] First six chapters of a projected twenty-four chapter treatise describing the current state-of-the-art in statistical computing. After a general introduction, covers basic numerical methods, numerical linear algebra, nonlinear statistical methods, numerical integration and approximation, and smoothing and density estimation. Emphasizes computational methods which are of general importance in statistics. RSK

**Statistics, P, L.** *Discriminant Analysis and Clustering.* Board on Mathematical Sciences. National Academy Pr, 1988, x + 105 pp, (P). A state-of-the-art survey of statistical classification techniques, covering cases where the groups are known (discriminant analysis) as well as where they are unknown (cluster analysis). Three main chapters survey methods, theory, and common software implementations (P-STAT, BMDP, SASS, SPSS, etc.). (Two intended chapters—on applications and on pedagogical aspects—are not included.) Intended to help focus current research on areas of greatest need. LAS

**Statistics, T(14-17: 1), S, P, L.** *Decision Analysis: A Bayesian Approach.* J.Q. Smith. Chapman & Hall, 1988, x + 138 pp, $25 (P); $47.50. [ISBN: 0-412-27520-1; 0-412-27510-4] Decision trees, utilities, subjective probabilities and their measurement, influence diagrams, group decision making, Bayesian statistics, and Bayesian estimation. Presupposes a course in statistics. The price seems high. FLW

**Statistics, P.** *Parameter Estimation and Hypothesis Testing in Linear Models.* Karl-Rudolf Koch. Springer-Verlag, 1988, xvi + 378 pp, $49.50 (P). [ISBN: 0-387-18840-1] Modified translation of the author's 1987 Second German Edition, written especially for engineers. Terse, primarily theoretical presentation, half devoted to prerequisite matrix algebra and probability theory, and half to multidimensional methods of estimating parameters, testing hypotheses, and estimating intervals. RSK

**Statistics, P.** *Regression Analysis and Empirical Processes.* S.A. van de Geer. CWI Tract, V. 45. Math Centrum, 1988, 161 pp, Dfl. 25.30 (P). [ISBN: 90-6196-330-3] Technical monograph concerned with least squares estimation of a regression function. Uses empirical process theory (uniform laws of large numbers and uniform central limit theorems) to describe asymptotic behavior of the estimators. General theory is illustrated using two-phase regression models. RSK

**Statistics, P*.** *Estimation of Variance Components and Applications.* C.R. Rao, J. Kleffe. Stat. &

Prob., V. 3. North-Holland (US Distr: Elsevier Science), 1988, xiii + 370 pp, $78. [ISBN: 0-444-70023-4] Covers essentially all known results on estimation of variance components through MINQE (minimum norm quadratic estimation), IMINQE (iterated MINQE), ML (maximum likelihood), and MML (marginal ML) methods. Also includes results on asymptotic distributions of estimators. Good set of references. Note price. RSK

**Statistics, S(16-18), P*, L.** *Sensitivity Analysis in Linear Regression.* Samprit Chatterjee, Ali S. Hadi. Prob. & Math. Stat. Wiley, 1988, xiv + 315 pp, $29.95. [ISBN: 0-471-82216-7] Examines the factors (variables used, observations, model assumptions) that determine a least squares fit, and studies the sensitivity of the fit to these factors. Good summary of known results and procedures, illustrated by numerous examples and diagnostic plots. RSK

**Statistics, P*.** *The Statistical Theory of Linear Systems.* E.J. Hannan, Manfred Deistler. Prob. & Math. Stat. Wiley, 1988, xiii + 380 pp, $42.95. [ISBN: 0-471-80777-X] Integrates recent developments in the theory and applications of system identification, including advances in signal processing and time-series analysis, in understanding the algebraic and topological structure of linear dynamic systems, and in asymptotic theory for estimation procedures. Concerned with linear dynamic systems evolving in discrete time, principally the stationary case and the case of noise-free inputs. Good set of references. RSK

**Statistics, P*.** *Analysis of Means in Some Non-standard Situations.* J.B. Dijkstra. CWI Tract, V. 47. Math Centrum, 1988, 138 pp, Dfl. 21.60 (P). [ISBN: 90-6196-347-8] Compares the effectiveness of different methods of testing hypotheses about the equality of several means, including standard, non-parametric and adaptive procedures, when assumptions for a classical test are not fulfilled. Particularly concerned with normal distributions having heterogeneity or extreme outliers, including the effects of these on multiple comparison procedures. RSK

**Statistics, P.** *Statistical Models and Analysis in Auditing.* National Research Council, 1988, xi + 91 pp, (P). A survey of statistical methods appropriate for auditing situations in which data distributions arise from nonstandard mixtures of distributions, one of which is degenerate (always zero). Such distributions arise in tax audits, in measuring rainfall, and in studies of smoking: in each some fraction of the data is zero (error-free, no precipitation, never smoked) while another part varies in a standard manner. LAS

**Statistics, P.** *Lecture Notes in Statistics-44: The Theory and Applications of Statistical Inference Functions.* D.L. McLeish, Christopher G. Small. Springer-Verlag, 1988, vi + 124 pp, $25 (P). [ISBN: 0-387-96720-6] Theoretical monograph extending standard concepts of ancillarity, sufficiency and completeness to classes of estimating functions (inference functions), with applications to estimation, censoring, robustness and inferential separation of parameters. RSK

**Elementary Computer Science, T(13-14: 1).** *Computer Organization and Architecture with Business Applications.* James Adair. Scott Foresman, 1988, xviii + 381 pp, $23.96. [ISBN: 0-673-18834-5] An elementary textbook in computer organization slanted towards business computer systems. Presents many of the traditional topics but not in as much detail as a similar textbook designed for use in a computer science course. AO

**Elementary Computer Science, T(13-14: 1, 2).** *Data Structures, Algorithms, and Program Style Using C.* James F. Korsh, Leonard J. Garrett. PWS-Kent, 1988, xvi + 590 pp, $28. [ISBN: 0-87150-099-X] Designed to be used in a second programming course as well as in an introductory data structures course. Good software engineering practices are emphasised throughout. AO

**Programming, T(14), S.** *Advanced Turbo Pascal Version 4.* Herbert Schildt. Osborne McGraw-Hill, 1988, xii + 416 pp, $21.95 (P). [ISBN: 0-007-881355-7] Turbo Pascal Version 4 is a powerful new programming system available for IBM PC and PC-compatible systems. This text introduces the intermediate Pascal programmer to a number of the new and very powerful features of Turbo Pascal, including the Database Toolbox and the Graphics Toolbox. It also demonstrates the capabilities of Pascal in such advanced areas as statistics, encryption, and language parsing. MGS

**Programming, S(13-14), L.** *Assembly Programming and the 8086 Microprocessor.* D.S. Jones. Oxford U Pr, 1988, vii + 203 pp, $42.50; $19.95 (P). [ISBN: 0-19-853743-3] A gentle introduction to assembly language programming on Intel 8086/8088, written by a Scottish mathematics professor "for other amateurs." The author's non-professional perspective pays off for the self-learner in sensible sequencing, helpful motivation; however, beware implicit dependence on Intel assembler, PC BIOS (e.g., CSEG, DSEG directives apparently unavailable on IBM assembler). About 100 exercises. Refreshing British tone. RB

**Programming, T(13).** *Short Fortran.* Robert P. Webber. Kendall/Hunt, 1988, 138 pp, $14.95 (P). [ISBN: 0-8403-4624-7] For self-study or a short course. Assumes knowledge of another high-level language. FORTRAN IV-based, but gives some details of WATFIV and FORTRAN 77. Drill exercises and short programming problems. DFA

**Programming, S, P, L.** *Using Turbo Pascal Version 4.* Steve Wood. Osborne McGraw-Hill, 1988, xii + 546 pp, $19.95 (P). [ISBN: 0-07-881356-5] For students with programming experience (not necessarily in Pascal) and even experienced Pascal programmers. Part I reviews Pascal and explains the language enhancements and software development environment of *Turbo Pascal Version 4.* Part II develops applications using general-purpose building blocks. Excellent reference; not a text. An appendix contrasts Versions 4 and 3. DFA

**Programming, T(14-15), S.** *Turbo C: The Complete Reference.* Herbert Schildt. Osborne McGraw-Hill, 1988, xviii + 908 pp, $24.95 (P). [ISBN: 0-07-881346-8] C is a system's implementation language used to build computer software. Turbo C is one of the most popular implementations of this language. The text is a thorough and complete introduction to this programming language for the person with no prior knowledge of it. The book includes a description of the language as well as the support tools provided with the system. MGS

**Programming, T(13: 1).** *Structured Programming Using Turbo BASIC.* Wade Ellis, Jr., Ed Lodi. Academic Pr, 1988, xvii + 337 pp, $23.95 (P). [ISBN: 0-12-237460-6] No prerequisites. Suitable for the first course in programming or for self-study. Many examples, most from business and elementary mathematics. Review questions, programming problems. Final chapters concern advanced graphics and simulation. Attractive; appears easy to use. DFA

**Programming, T(13-14), S.** *Advanced Structured Basic: File Processing with the IBM/PC.* James Payne. PWS-Kent, 1988, x + 290 pp, $25.50 (P). [ISBN: 0-534-91872-7] Assumes knowledge of BASIC, DOS, and IBM PC machines (there is a short appendix covering DOS and PC's). Chapter 2 covers (structured) flowcharts and pseudocode, which are used consistently throughout the text. Each chapter has an overview, a summary and review, key terms, and exercises. Sequential files, random files, and indexed sequential files are included, as well as a chapter on graphics. RSF

**Programming, S(13-15), P, L.** *The C Programming Language, Second Edition.* Brian W. Kernighan, Dennis M. Ritchie. Prentice-Hall, 1988, xii + 272 pp, (P). [ISBN: 0-13-110362-8] This *Second Edition* of "K&R," the classic reference to C, is based on the draft-proposed ANSI C standard. Besides incorporating ANSI type checking and standard library, there are some exposition changes which will significantly assist C learners, including alternate choice of examples and helpful diagrams illustrating pointers, structures, etc. Appendix of changes between editions; *First Edition* (TR, May 1979) will remain in print as reference for original language. RB

**Languages, P.** *Program Correctness over Abstract Data Types, with Error-State Semantics.* J.V. Tucker, J.I. Zucker. CWI Mono., V. 6. North-Holland (US Distr: Elsevier Science), 1988, viii + 212 pp, $58. [ISBN: 0-444-70340-3] Tools for proving correctness in an abstract setting where use of an uninitialized variable is an error. These include classes of many-sorted structures, weak second order assertion languages, generalisation of computable function theory to the many-sorted structures. Discusses Church-Turing thesis extensions. DFA

**Languages, P.** *Lecture Notes in Computer Science-300: ESOP '88.* Ed: H. Ganzinger. Springer-Verlag, 1988, vi + 381 pp, $30.30 (P). [ISBN: 0-387-19027-9] 24 papers presented at the Second European Symposium on Programming held in Nancy, France, March 21-24, 1988, the theme of which was the design, specification, and implementation of programming languages and systems. Also describes some software systems exhibited at the conference. DFA

**Languages, S?(13).** *Advanced C, Second Edition.* Herbert Schildt. Osborne McGraw-Hill, 1988, xi + 403 pp, $21.95 (P). [ISBN: 0-07-881348-4] This book is *not* a textbook on C language programming. A collection of sample programs, it is probably best suited for the hobbyist who wants to see examples of C language programs. AO

**Algorithms, T(15: 1).** *Files and Data Structures with COBOL.* James Mensching. Scott Foresman, 1988, 368 pp, $31.95. [ISBN: 0-673-18608-3] For the programming student interested in business applications. Assumes familiarity with COBOL and basic understanding of computer hardware and software. File processing treatment discusses specific operating system and programming environments; examples and problems are in COBOL. Treats data structures more theoretically, using pseudocode. Discusses database management systems and issues. DFA

**Algorithms, T(16-17: 1), L.** *Algorithm Design: A Recursion Transformation Framework.* Marvin C. Paull. Wiley, 1988, xiv + 490 pp, $44.95. [ISBN: 0-471-81688-4] Presents a general framework for algorithm design starting from a recursive definition of the function to be evaluated. Also discusses the problem of formulating a recursive definition of a function starting from an initial definition that is not recursive. Emphasizes the general principles that underlie the design of efficient algorithms rather than the design of data structures. AO

**Algorithms, P.** *Algorithmic Methods in Algebra and Number Theory.* Ed: Michael Pohst. Academic Pr, 1987, 135 pp, $12.50 (P). [ISBN: 0-12-559190-X] A collection of fourteen papers. This special issue of the *Journal of Symbolic Computation* is dedicated to Hans Zassenhaus on the occasion of his 75th birthday. Includes a bibliography of the honoree's work. CEC

**Algorithms, P, L.** *The Rapid Evaluation of Potential Fields in Particle Systems.* Leslie Greengard. ACM Disting. Dissertations. MIT Pr, 1988, xi + 91 pp, $25. [ISBN: 0-262-07110-X] Simulating inverse-square (Coulombic or gravitational) interactions between pairs of $N$ particles would seem to require $O(N^2)$ computations. The author has lowered it to $O(N)$ using multi-pole expansions. This book shows how. BC

**Algorithms, S(18), P, L.** *Resource Allocation Problems: Algorithmic Approaches.* Toshihide Ibaraki, Naoki Katoh. MIT Pr, 1988, xiv + 229 pp, $37.50. [ISBN: 0-262-09027-9] The "resource allocation problem" is a mathematical programming problem which seeks to optimize an objective function by effectively allocating a single resource. The sole constraint represents the availability of the resource. Decision variables may be continuous or discrete. The objective function might be convex and/or separable. This book summarises algorithms developed over the past

thirty years to solve various instances of the problem. No exercises. SM

**Computer Systems, P.** *Lecture Notes in Computer Science-297: Supercomputing.* Ed: E.N. Houstis, T.S. Papatheodorou, C.D. Polychronopoulos. Springer-Verlag, 1988, x + 1093 pp, $75 (P). [ISBN: 0-387-18991-2] Proceedings of the first international conference on supercomputing, Athens, Greece, June 1987. 63 papers on parallel processing, parallel architectures, software environments for parallel machines, compilers and restructuring techniques, problem mapping and scheduling, parallel numeric methods, VLSI, dataflow and array processors, and algorithms, architecture, and performance. RB

**Computer Systems, S(13).** *UNIX: The Complete Reference: System V, Release 3.* Stephen Coffin. Osborne McGraw-Hill, 1988, 704 pp, $24.95 (P). [ISBN: 0-07-881299-2] An introduction to the UNIX operating system (specifically, the System V, Release 3 version) for users without previous experience with UNIX. It provides an overview of the capabilities and special features of UNIX system software in a manner that is easy to read and understand. A good introduction to UNIX for new users, but, despite its title, not a reference manual. AO

**Computer Graphics, P, L?** *CGM and CGI: Metafile and Interface Standards for Computer Graphics.* David B. Arnold, Peter R. Bono. Springer-Verlag, 1988, xxi + 279 pp, $39.50. [ISBN: 0-387-18950-5] Description and comparison with other standardization efforts of two ISO computer graphics standardization projects for the Virtual Device Interface: Computer Graphics Interface, for standardization of functional and syntactic specifications for the exchange of device independent data and control information, and Computer Graphics Metafile, for file formats suitable for storage and retrieval of device independent picture descriptions. RM

**Computer Graphics, P, L.** *Advances in Computer Graphics III.* Ed: M.M. de Ruiter. EurographicSeminars. Springer-Verlag, 1988, ix + 322 pp, $72. [ISBN: 0-387-18788-X] Proceedings of Eurographics '87 conference, Amsterdam, August 1987. Eight surveys by Dutch, German, English, Danish authors: VLSI support for solid modelling; user interface management systems; object-oriented graphics; computer graphics in art; fractals; 3-dimensional geometric modelling techniques; CAD interface specifications; desk-top publishing. RB

**Theory of Computation, P.** *Lecture Notes in Computer Science-299: CAAP '88.* Ed: M. Dauchet, M. Nivat. Springer-Verlag, 1988, viii + 304 pp, $27.30 (P). [ISBN: 0-387-19021-X] Proceedings of the 13th Colloquium on Trees in Algebra and Programming, with papers on trees, tree grammars and complexity, abstract data types and term rewriting, algebraic specifications, logics, parallelism, and concurrency. RM

**Theory of Computation, P.** *Fully Abstract Models of Programming Languages.* Allen Stoughton. Pitman (US Distr: Wiley), 1988, 123 pp, $19.95 (P). [ISBN: 0-470-21041-9] This is a monograph of a theory of denotational semantics of programming languages. It is extraordinarily complex and is intended only for the serious researcher in abstract semantics of formal languages. It requires a significant background in set theory, category theory, abstract algebra, and theoretical computer science. Approach with extreme caution! MGS

**Theory of Computation, P.** *Understanding Z: A Specification Language and its Formal Semantics.* J.M. Spivey. Tracts in Theoretical Comput. Sci., V. 3. Cambridge U Pr, 1988, viii + 131 pp, $27.95. [ISBN: 0-521-33429-2] The $Z$ notation expresses formal specifications of practical computing systems, based on typed set theory and "schemas." This monograph gives formal semantics for $Z$ in terms of mathematical logic, via denotational semantics. Applications of $Z$, comparison with other specification methods. Not intended as introduction to $Z$ (which is under development at Oxford). RB

**Artificial Intelligence, S(16-17).** *Expert System Applications.* Ed: L. Bolc, M.J. Coombs. Symb. Computat. Springer-Verlag, 1988, ix + 471 pp, $59.50. [ISBN: 0-387-18722-7] Expert systems are computer programs which demonstrate decision-making abilities in a specific application area as good as or better than a human expert. This text contains papers describing nine expert systems research projects from the U.S., Japan, and France in a range of areas including medical diagnosis, the design of electrical circuits, and analyzing x-ray images. MGS

**Artificial Intelligence, S(16-18), P, L.** *What Every Engineer Should Know About Artificial Intelligence.* William A. Taylor. MIT Pr, 1988, xi + 331 pp, $25. [ISBN: 0-262-20069-4] Guided tour of artificial intelligence: history, successes and failures, current research issues, artificial intelligence's influence on current practice in computer science and engineering. Nice overview for technical people who want a sense of the subject, though not very deep. Good annotated bibliography for further study. RM

**Artificial Intelligence, S, P, L*.** *Computer Games II.* Ed: David N.L. Levy. Springer-Verlag, 1988, xv + 546 pp, $59. [ISBN: 0-387-96609-9] Second of a two-volume anthology (*Volume I*, TR, December 1988) of seminal papers and theses on computer methods for games of strategy. This volume features Go—the most difficult game of strategy—with twelve papers occupying nearly half the volume. Other games include Bridge, Othello, Poker, and Dominoes. Concludes with a comprehensive bibliography covering both volumes. LAS

**Artificial Intelligence, P.** *Machine Vision: Algorithms, Architectures, and Systems.* Ed: Herbert Freeman. Perspectives in Comput., V. 20. Academic Pr, 1988, x + 315 pp, $39.95. [ISBN: 0-12-266720-4] Proceedings of a workshop at Rutgers University, April 1987. Fourteen papers on specific computer architectures and criteria, industrial machine vision, bottlenecks to effective applications, algorithms for machine inference of visual features, future research

directions. RB

**Computer Science, P.** *Lecture Notes in Computer Science-307: Applicable Algebra, Error-Correcting Codes, Combinatorics and Computer Algebra.* Ed: Th. Beth, M. Clausen. Springer-Verlag, 1988, vi + 215 pp, $21.80 (P). [ISBN: 0-387-19200-X] Proceedings of the conference AAECC-4, Karlsruhe, West Germany, September 1986. LC

**Computer Science, P.** *Lecture Notes in Computer Science-303: Advances in Database Technology—EDBT '88.* Ed: J.W. Schmidt, S. Ceri, M. Missikoff. Springer-Verlag, 1988, x + 620 pp, $48.50 (P). [ISBN: 0-387-19074-0] Proceedings of the International Conference on Extending Database Technology held March 14-18, 1988 in Venice, Italy. The purpose of the conference was to share information about research that extends the scope of database technology: extended semantics, new architectures and support systems, and applications such as heterogeneous and multimedia databases. AO

**Computer Science, P.** *Mathematical Aspects of Scientific Software.* Ed: J.R. Rice. IMA, V. 14. Springer-Verlag, 1988, xi + 208 pp, $21. [ISBN: 0-387-96706-0] Partial proceedings of a workshop in 1986-87 IMA (Institute for Mathematics and Its Applications) program in Scientific Computation. Eight papers on the influence of scientific programming on mathematics (not vice versa) concerning parallel computations, geometric computations, general use of symbolic computation packages and extended packages, performance measurement of computational tools. RB

**Applications, P.** *Transactions of the Fifth Army Conference on Applied Mathematics and Computing.* US Army Research Office (PO Box 1211, Research Triangle Park, NC 27709), 1988, xx + 846 pp, (P). Forty-three invited and contributed papers from the conference held at West Point, New York, from June 15-18, 1987. AO

**Applications, P, L.** *Population System Control.* Jian Song, Jingyuan Yu. Springer-Verlag, 1988, xi + 286 pp, $49.90. [ISBN: 0-387-18288-8] A scholarly mixture of sociology and classical mathematical methods of demography, focused primarily on China with ample data on world population trends. Mathematical focus is on stability theory—on the critical value of fertility beyond which population explodes without limit—and on strategies of optimal control theory which might nudge nations towards a stable, zero-growth population pattern. LAS

**Applications, P.** *Selected Papers on the Teaching of Mathematics as a Service Subject.* Ed: R.R. Clements, P. Lauginie, E. de Turckheim. Springer-Verlag, 1988, 181 pp, $35 (P). [ISBN: 0-387-82056-6] A supplement to the ICMI Study *Mathematics as a Service Subject* (TR, October 1988) derived from a 1987 workshop in Udine, Italy. Twenty papers describing issues in teaching mathematics to nonmathematicians from more than a dozen different countries. A good sampler of strategies, customs, problems, and accomplishments in diverse cultural settings. LAS

**Applications (Engineering), T(17-18: 2), L.** *Numerical Modeling in Science and Engineering.* Myron B. Allen III, Ismael Herrera, George F. Pinder. Wiley, 1988, x + 418 pp, $39.95. [ISBN: 0-471-80635-8] For engineers, scientists, applied mathematicians. Basic continuum mechanics and numerical methods for partial differential equations; physical, mathematical, and numerical aspects of steady-state systems, dissipative systems, and nondissipative systems; modeling of higher-order equations, nonlinear equations, coupled systems of equations. Examples with illustrations, problems, references. DFA

**Applications (Fluid Mechanics), P, L.** *Mathematical Models in Environmental Problems.* G.I. Marchuk. Studies in Math. & Its Applic., V. 16. North-Holland (US Distr: Elsevier Science), 1986, 217 pp, $51.75. [ISBN: 0-444-87965-X] Deals with estimation of atmospheric pollution and the contamination of an underlying surface with passive and active pollutants. Rates of deposition, advective transfer, and diffusion are covered. Goal is to optimize location of new industrial facilities. SM

**Applications (Fluid Mechanics), T(18: 1).** *Boundary Value Problems in Linear Viscoelasticity.* J.M. Golden, G.A.C. Graham. Springer-Verlag, 1988, xiv + 266 pp, $72.50. [ISBN: 0-387-18615-8] A graduate textbook. Exercises are interspersed throughout the body of the text. Types of problems covered include planar non-inertial contact and crack problems, three-dimensional contact problems, and planar inertial problems. SM

**Applications (Medicine), S*(12-14).** *Mathematics and Medicine: How Serious is the Injury?* William Sacco, *et al.* Contemp. Appl. Math. Janson, 1987, viii + 53 pp, $7.95 (P). [ISBN: 0-939765-06-3] Self-teaching module showing how mathematical indices which measure the severity of patient illness or injury can be defined, constructed, and applied. Applications include triage (where to take a patient), sorting of patients, patient tracking, and evaluation of patient care. Part of a series "designed to introduce secondary school or college students to exciting mathematical topics that have interesting and important applications." RSK

**Applications (Physical Science), P.** *Quantitative Analysis of Mineral and Energy Resources.* Ed: C.F. Chung, A.G. Fabbri, R. Sinding-Larsen. NATO ASI Ser. C: V. 223. D Reidel (US Distr: Kluwer Academic), 1988, xviii + 738 pp, $139. [ISBN: 90-277-2635-3] Contains 40 papers presented at the NATO Advanced Study Institute on "Statistical Treatments for Estimation of Mineral and Energy Resources" held at Il Ciocca, Italy on June 22-July 4, 1986. Concludes with five reports on current problems and future developments in various specialties germane to geostatistics. SM

**Applications (Physics), P.** *Lecture Notes in Mathematics-1303: Quantum Probability and Applications III.* Ed: L. Accardi, W. von Waldenfels.

Springer-Verlag, 1988, vi + 373 pp, $34.80 (P). [ISBN: 0-387-18919-X] Twenty-seven papers from a conference on quantum probability. BC

**Applications (Physics), P.** *Theory and Applications of Inverse Problems.* Ed: H. Haario. Pitman Res. Notes in Math Ser., V. 167. Longman Scientific & Technical (US Distr: Wiley), 1988, 159 pp, $46.95 (P). [ISBN: 0-582-01479-4] Papers presented at a symposium of mathematicians and physicists at the University of Helsinki in 1985 on the theme of "inverse problems:" the inverse scattering problem of quantum physics, determination of the shape of a sound scattering obstacle from a knowledge of the far field scattered acoustic wave, the inversion of the Laplace transform, etc. AWR

**Applications (Physics), P.** *Introduction to Supersymmetry.* Peter G.O. Freund. Cambridge U Pr, 1986, x + 152 pp, $14.95 (P). [ISBN: 0-521-35675-X] Brief introductory text focusing on basic ideas and techniques. Supersymmetry, at the very forefront of theoretical physics, has taken on a life of its own growing beyond the ability of experimentalists to verify it. "...we are faced here, with one of those rare instances, where the mathematicians, in all their wisdom, have overlooked a beautiful structure, ...and have come to appreciate it only at the demands of physicists." A compliment, in a way. (1986 hardcover text, TR, November 1987.) MR

**Applications (Physics), P.** *Proceedings Seminar 1984-1986: Mathematical Structures in Field Theories, Volume 1.* P.J.M. Bongaarts, E.A. de Kerf, P.H.M. Kersten. CWI Syllabus, V. 16. Math Centrum, 1988, iii + 137 pp, Dfl. 21.60 (P). [ISBN: 90-6196-345-1] Contains three papers from seminars at the University of Amsterdam from 1984-86. First paper discusses the axiomatics of free quantum systems; the second continues with a discussion of the physical aspects of those fields. The last paper studies infinitesimal symmetries using the local jet bundle. MR

**Applications (Physics), P.** *Lecture Notes on Particle Systems and Percolation.* Richard Durrett. Wadsworth, 1988, viii + 335 pp, $39.95. [ISBN: 0-534-09462-7] An almost informal introduction to percolation theory, divided into four acts (authors term). Act 1 centers on simple particle systems such as voter models. Act 2 covers discrete time models. Act 3 covers percolation and first passage percolation. Includes a chapter on fractals. Act 4 is more technical than previous acts and gives recent results in the subject. Text is clear and ideas are well-motivated by applications and examples. A disk containing computer simulations of examples treated in the book is available from the publisher. MR

**Applications (Physics), T(17-18: 1, 2), S*, P, L.** *Introduction to Superstrings.* Michio Kaku. Grad. Texts in Contemp. Physics. Springer-Verlag, 1988, xvi + 568 pp, $49.95. [ISBN: 0-387-96700-1] An ambitious attempt to unify a rapidly evolving subject which "often seems like a confused jumble of folklore, random rules of thumb, and intuition." Opens by quoting Niels Bohr: "We are all agreed that your theory is crazy. The question which divides us is whether it is crazy enough." BC

**Applications (Physics), S(13-14), L*.** *Modelling with Projectiles.* Derek Hart, Tony Croft. Math. & Its Applic. Halsted Pr, 1988, 152 pp, $39.95. [ISBN: 0-470-21085-0] An in-depth analysis of projectile motion from golf balls to javelins. Uses basic techniques of vector calculus to develop equations of motion in resisting media; employs the enveloping parabola (that contains all trajectories with the same initial velocity) as well as computer programs (for numerical estimates) as tools to solve a variety of problems. LAS

**Applications (Physics), T(17-18: 1, 2), S, P, L.** *Mathematical Elasticity, Volume I: Three-Dimensional Elasticity.* Philippe G. Ciarlet. Stud. in Math. & Its Applic., V. 20. North-Holland (US Distr: Elsevier Science), 1988, xl + 451 pp, $107.25. [ISBN: 0-444-70259-8] Graduate-level introduction to current theories of three-dimensional elasticity (static only—time-dependent elasticity is still poorly understood), emphasizing nonlinear models and recent existence results. BC

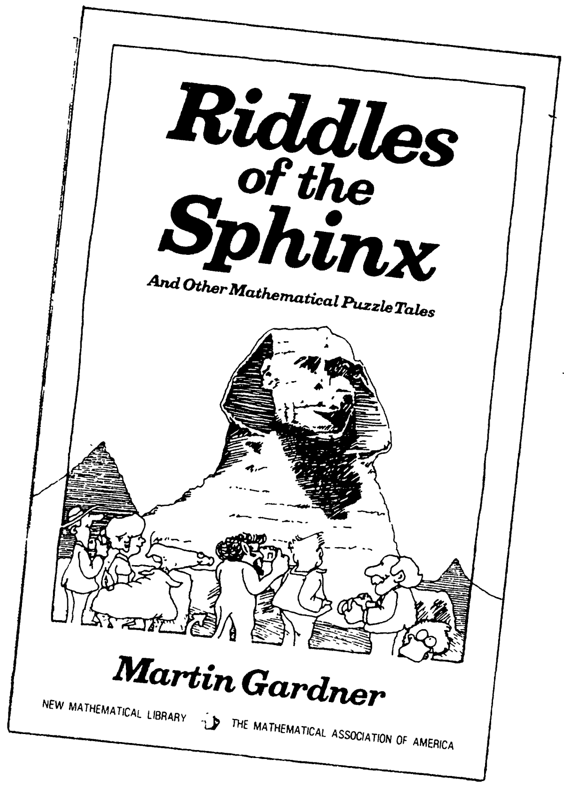
**Applications (Social Science), P, L*.** *Making Multicandidate Elections More Democratic.* Samuel Merrill, III. Princeton U Pr, 1988, xix + 149 pp, $29.50. [ISBN: 0-691-07770-3] A systematic analysis of the extent to which various election systems enhance desirable criteria for social choice systems. Based on Monte-Carlo simulations, studies of actual cases, and theoretical analyses (in appendices). One over-riding conclusion: single-vote plurality—the most widely-used election system in America—is the weakest among all the options. LAS

### Reviewers

RJA: Richard J. Allen, St. Olaf; DFA: David F. Appleyard, Carleton; SB: Steve Benson, St. Olaf; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; CE: Christopher Ennis, Carleton; RSF: Robert S. Fisk, St. Olaf; SG: Steven Galovich, Carleton; GG: George Gilbert, St. Olaf; BH: Bruce Hanson, St. Olaf; RH: Rhonda Hatcher, St. Olaf; TH: Timothy Hesterberg, St. Olaf; PH: Paul Humke, St. Olaf; JJ: Jason Jones, St. Olaf; RBK: Roger B. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; GN: Gail Nelson, Carleton; AO: Arnold Ostebee, St. Olaf; SP: Samuel Patterson, Carleton; MR: Matthew Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; GMS: G. Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MS: Myriam Steinback, Macalester; ES: Elisabeth Strouse, Macalester; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; LW: Lynne Walling, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

# Riddles of the Sphinx

## and other mathematical puzzle tales.




by Martin Gardner
Volume 32 in the New Mathematical Library
184 pp., Paper. ISBN-0-88285-632-8
List: $14.50    MAA Member: $12.50

Martin Gardner has charmed readers for over fifty years with his delightful books and articles on science. He is best known for the popular column, *"Mathematical Games,"* which appeared in *Scientific American* for twenty-five years. Generations of scientists and mathematicians have been inspired by his writing and the MAA is proud to include his name in its list of authors.

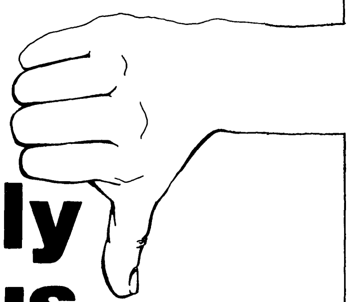
This book was drawn from Gardner's column in *Isaac Asimov's Science Fiction Magazine.* The riddles presented here incorporate the responses of his initial readers, along with additions suggested by the editors of the New Mathematical Library. Each chapter (riddle) poses a problem answered in the First Answers section. The solution in turn raises another problem that is solved in the Second Answers section. This may suggest a third question and in several instances there is a fourth. Gardner draws us from questions to answers always presenting us with new riddles—some as yet unanswered. There are 125 different pieces altogether.

Solving these riddles is not simply a matter of logic and calculation, although these play a role. Luck and inspiration are factors as well, so beginners and experts alike may profitably exercise their wits on Gardner's problems, whose subjects range from geometry to word play to questions relating to physics and geology.

We guarantee that you will solve some of the riddles, be stumped by others, and be amused by almost all of the stories and settings that Gardner has devised to raise these questions.

Order from:
**The Mathematical Association of America**
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

# Toward a Lean and Lively Calculus



**Report of the Conference/Workshop to Develop Curriculum and Teaching Methods for Calculus at the College Level, Ronald G. Douglas, Editor.**
**MAA Notes #6**
249 pp., 1987, Paperbound, ISBN-0-88385-056-7
Catalog Number - NTE-07
Price: $12.50

Should calculus be taught differently? Can it? Common wisdom says "no"—which topics are taught, and when, are dictated by the logic of the subject and by client department. The surprising answer from a four-day Sloan Foundation-sponsored conference on calculus instruction, chaired by Ronald Douglas, is that significant change is possible, desirable, and necessary. Meeting at Tulane University in New Orleans in January, 1986, a diverse and sometimes contentious group of twenty-five faculty, university and foundation administrators, and scientists from client departments, put aside their differences to call for a leaner, livelier, more contemporary course, more sharply focused on calculus's central ideas and on its role as the language of science.

This volume contains the results of that conference and the papers presented to the conferees. These are certain to be the point of departure and basis for efforts to strengthen and reshape calculus in the next decade.

Order from
The Mathematical Association
of America
1529 Eighteenth St., NW
Washington, DC 20036

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 96, Number 2 February 1989

# Contents

(ISSN 0002-9890)

ARTICLES

Patterns in Linear Algebra . . . . . . . . . GILBERT STRANG 105

Chebyshev's Inequality and Natural Density . . . . . . . CURTIS N. COOPER AND ROBERT E. KENNEDY 118

EDITOR'S CORNER

The New Mersenne Conjecture P. T. BATEMAN, J. L. SELFRIDGE, AND S. S. WAGSTAFF, JR. 125

UNSOLVED PROBLEMS

Polynomials All of Whose Derivatives Have Integer Roots . . . . C. E. CARROLL 129

NOTES

Generalizing the Formula for Areas of Polygons to Moments . . . . S. F. BOCKMAN 131

An Elementary Test for the Galois Group of a Quartic Polynomial . . . . . . LUISE-CHARLOTTE KAPPE AND BETTE WARREN 133

Comparing the Spectral Radii of Two Nonnegative Matrices . . . . . . . R. B. BAPAT 137

THE TEACHING OF MATHEMATICS

From Calculus to Number Theory . . . . . . . . JAMES DUEMMEL 140

A Note on the Row-Reduction Algorithm . . . . . . . . CHIH-HAN SAH 143

On Markov Processes in Elementary Mathematics Courses . . JOHN T. BALDWIN 147

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions . . . . . . . . 154

Advanced Problems and Solutions . . . . . . . . 165

REVIEWS

Littlewood's Miscellany by Béla Bollobás . . . . . . . . RALPH P. BOAS 167

Forever Undecided: A Puzzle Guide to Gödel by Raymond Smullyan . . . . . . . . CRAIG SMORYNSKI 169

TELEGRAPHIC REVIEWS . . . . . . . . 173

## NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Dept. of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to ANITA E. SOLOW, Dept. of Mathematics, Grinnell College, Grinnell, IA 50112; in geometry or topology to DENNIS DETURCK, Dept. of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems (three copies) and solutions (two copies), both elementary and advanced to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, University of Calgary, Alberta, Canada T2N 1N4.




Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: MS. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $26 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $29. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $90 per year.





Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Assocation of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

# Patterns in Linear Algebra

GILBERT STRANG, *Massachusetts Institute of Technology*

GILBERT STRANG received his Ph.D. in 1959 under the guidance of Peter Henrici at UCLA. His teaching at MIT led to the textbook *Linear Algebra and Its Applications*, and in preparing the third edition the matrices in this paper turned up as valuable examples. They also illustrate the "$A^TCA$ pattern" that dominates so many applications and is developed in the new text *Introduction to Applied Mathematics*. The author's current research involves Toeplitz matrices and the Fast Fourier Transform—in other words, more matrices.

This article is about two remarkable families of matrices. As examples in linear algebra, they are just about perfect. They illustrate the central ideas of elimination and diagonalization and orthogonalization (and the factorizations into $LDL^T$ and $Q\Lambda Q^T$ and $QR$) with numbers that make you smile. Even the inverses come out right. The determinants are connected to the spanning trees of a graph, and they grow quickly for one family; the condition number grows more quickly for the other. But the real significance of these matrices goes beyond the patterns that appear when you compute with them. It is the *source* of the matrices that makes them important, and that is the main point of the paper.

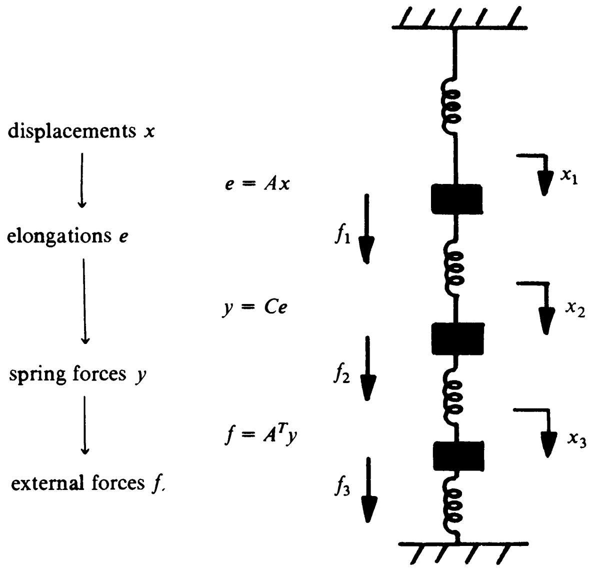
A matrix can be just a record of information or (like the registers of a von Neumann computer) it can contain instructions. It can express a law of physics or mechanics or graph theory. In that case the matrix represents a linear operator; *it can act*. For a continuous problem it becomes a differential or an integral operator; our first family approaches $-d^2/dx^2$ (and the second is not so clear). The key to everything is the framework from which the matrices are constructed—out of simpler matrices.

The starting point for the first family is a line of springs and masses. I have to admit that the end result is extremely straightforward, and you may be tempted to go right past the framework that gets us there. We will reach the symmetric matrices $T$ with three diagonals, containing $-1$ and $2$ and $-1$. They are "second-difference matrices," but they are built out of first differences and that pattern is worth seeing.

The goal is to connect the *displacements* $x_1, x_2, x_3$—the movements of the three masses from their neutral positions—to the *external forces* $f_1, f_2, f_3$ that produce these displacements. The direct path is $Tx = f$. The true path is through the springs! It is the *elongations* $e_1, e_2, e_3, e_4$ of the four springs that come from the displacements, and it is the *spring forces* $y_1, y_2, y_3, y_4$ that balance $f$. The gap between $x$ and $f$ is filled by $e$ and $y$ (see p. 106).

Each step is common sense more than physics. The first spring is stretched by $x_1$; the second is stretched by $x_2 - x_1$ (it would be unstretched if those two displacements were equal); the four elongations come from differences in the three displacements:

$$\begin{aligned} e_1 &= x_1 \\ e_2 &= x_2 - x_1 \\ e_3 &= x_3 - x_2 \\ e_4 &= -x_3 \end{aligned} \qquad \text{or} \qquad e = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = Ax.$$




In the step from $e$ to $y$, the matrix is square. In fact it is diagonal—each spring resists elongation by force, and in Hooke's law those quantities are proportional:

$$\begin{aligned} y_1 &= c_1 e_1 \\ y_2 &= c_2 e_2 \\ y_3 &= c_3 e_3 \\ y_4 &= c_4 e_4 \end{aligned} \quad \text{or} \quad y = \begin{bmatrix} c_1 & & & \\ & c_2 & & \\ & & c_3 & \\ & & & c_4 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} = Ce.$$

The final step is the key. It is a balance of forces, internal vs. external, and the matrix description brings out the crucial point. The first mass is pulled up by $y_1$, down by $y_2$, and the difference is matched by $f_1$:

$$\begin{aligned} f_1 &= y_1 - y_2 \\ f_2 &= y_2 - y_3 \\ f_3 &= y_3 - y_4 \end{aligned} \quad \text{or} \quad f = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = A^T y.$$

That matrix is the transpose of $A$. The laws of mechanics are symmetric for a conservative system at equilibrium, and the matrix that multiplied $x$ has reappeared —transposed—to multiply the dual unknown $y$. At one level this reflects a balance of external work (force times displacement at the masses) against internal work (force times elongation of the springs)

$$f^T x = y^T e \quad \text{or} \quad \left(A^T y\right)^T x = y^T (Ax).$$

We comment briefly on this framework of $A^T$, $C$, and $A$, before moving (in Section 1 below) to the numerical linear algebra. In mechanics as in mathematics, that last equation is the true definition of $A^T$. At a deeper level the system is governed by a *minimum principle*, and what we are seeing is the equilibrium equation for the position of minimum energy. $A^T C A x = f$ is an Euler-Lagrange

equation in the calculus of variations, and its symmetry is more fundamental than its linearity.†

The matrix $T$ compresses three steps into one: $f = A^Ty = A^TCe = A^TCAx$. *The coefficient matrix is* $T = A^TCA$, and matrix multiplication gives

$$T = \begin{bmatrix} c_1 + c_2 & -c_2 & 0 \\ -c_2 & c_2 + c_3 & -c_3 \\ 0 & -c_3 & c_3 + c_4 \end{bmatrix}.$$

Three properties of $T$ come directly from that construction. The matrix is *symmetric* because $A^TCA$ is always symmetric (if $C$ is). It is *tridiagonal* because the springs connect nearest neighbors. The third property is *positive definiteness*. Again there is a basic requirement on $C$—the coefficients $c_i$ must be positive, so $C$ itself is positive definite. Then linear algebra determines the key requirement on $A$:

*$A^TCA$ is positive definite if the columns of $A$ are independent.*

Algebraically, we must show that $x^TTx = (Ax)^TC(Ax) = e^TCe$ is never negative, and (for definiteness) that it is positive unless $x = 0$. Certainly the expression $e^TCe = c_1e_1^2 + \cdots + c_4e_4^2$ is not negative. It is zero only if $e = Ax = 0$. But with independent columns in $A$, that happens only when $x = 0$.

Mechanically, the masses cannot move without stretching a spring. There is no "rigid motion," and it is the boundary conditions $x_0 = x_4 = 0$ that have made the columns independent. If the supports were removed, the whole system could be shifted by $x = (1, 1, 1)$ with no stretching. $A^TCA$ would be singular. If only the lower support is removed, the boundary condition changes from $x_4 = 0$ (Dirichlet) to $y_4 = 0$ (Neumann). The columns of $A$ remain independent and $A^TCA$ is still positive definite.

Our matrix $T$ is certainly invertible; $Tx = f$ can be solved. To see clearly what numerical linear algebra does with such a matrix, we generalize to $n$ masses—but we also specialize to equal constants $c_i = 1$. Then $T = A^TA$ with $C = I$, and a typical matrix (in our first family) is

$$T_n = \begin{bmatrix} 2 & -1 & & \\ -1 & 2 & \cdot & \\ & \cdot & \cdot & -1 \\ & & -1 & 2 \end{bmatrix}.$$

By computing with $T = T_3$, the patterns will be clear for $T_n$.

### 1. Gaussian Elimination and $A = LDL^T$

Our first goal is to follow the elimination process on these matrices, and to recognize the underlying triangular factorization. Elimination starts with the pivot $d_1 = 2$; row 1 is multiplied by $\ell_{21} = -1/2$ and subtracted from row 2. Then the second pivot is $d_2 = 3/2$ and the multiplier is $\ell_{32} = -2/3$:

$$\begin{bmatrix} 2 & -1 & \\ -1 & 2 & -1 \\ & -1 & 2 \end{bmatrix} \to \begin{bmatrix} 2 & -1 & \\ 0 & 3/2 & -1 \\ & -1 & 2 \end{bmatrix} \to \begin{bmatrix} 2 & -1 & \\ & 3/2 & -1 \\ & & 4/3 \end{bmatrix}.$$

The third pivot is 4/3, on the diagonal of this upper triangular matrix $U$. *Gaussian*

†Equations $A^TC(Ax) = f$ with a nonlinear $C$ are developed in a companion paper [4] and in [3].

*elimination has factored $T$ into a product $LU$*, in which the lower triangular matrix $L$ contains the multipliers. It is displayed below, after one more step. $LU$ will reflect the symmetry of $T$ if we further divide the pivots from the rows of $U$, and reach the symmetric factorization $LDL^T$:

$$T=\begin{bmatrix}1&&\\-\frac{1}{2}&1&\\&-\frac{2}{3}&1\end{bmatrix}\begin{bmatrix}\frac{2}{1}&&\\&\frac{3}{2}&\\&&\frac{4}{3}\end{bmatrix}\begin{bmatrix}1&-\frac{1}{2}&\\&1&-\frac{2}{3}\\&&1\end{bmatrix}=LDL^T.$$

This factorization displays the three properties of $T$. The product is *symmetric* because $L$ appears with $L^T$, it is *tridiagonal* because $L$ is bidiagonal, and it is *positive definite* because the pivots are positive. For a symmetric matrix,

$$\textit{positive definiteness} \langle = \rangle \textit{positive pivots}.$$

The determinant (product of the pivots) is $\frac{2}{1}\,\frac{3}{2}\,\frac{4}{3}=4$. For $T_n$ it is $\frac{2}{1}\,\frac{3}{2}\cdots\frac{n+1}{n}=n+1$.

## 2. Orthogonalization and $A = QR$

Strictly speaking it is not recommended to work with $T = A^TA$ (although everybody does it). Instead of elimination on $T$, it is more stable to orthogonalize the columns of $A$. This is the Gram-Schmidt process—the bane of textbook authors, because the numbers are usually miserable. The orthogonalized vectors appear successively in the columns, and for this $A$ they are beautiful:

$$A=\begin{bmatrix}1&&\\-1&1&\\&-1&1\\&&-1\end{bmatrix}\to\begin{bmatrix}1&\frac{1}{2}&\\-1&\frac{1}{2}&\\&-1&1\\&&-1\end{bmatrix}\to\begin{bmatrix}1&\frac{1}{2}&\frac{1}{3}\\-1&\frac{1}{2}&\frac{1}{3}\\&-1&\frac{1}{3}\\&&-1\end{bmatrix}=B.$$

The next orthogonal column would be $1/4, 1/4, 1/4, 1/4, -1$, a pattern I have never seen before. All columns are perpendicular to the vector $(1, 1, \ldots, 1)$.

It is important to look more closely at the multipliers. Gram-Schmidt has multiplied column 1 by $-1/2$, and subtracted it to make the second column orthogonal. At the next step $-2/3$ of the new second column was subtracted from the third, to produce an orthogonal third column. *Those are the same multipliers that were used in elimination*! The steps of Gram-Schmidt, to orthogonalize the columns of $A$, correspond exactly to elimination steps on $T = A^TA$. If that is true, then the pivots of $T$ (which were 2, 3/2, 4/3) should also appear. You will recognize them as the squares of the lengths of the columns (for example $(1/2)^2 + (1/2)^2 + (-1)^2 = 3/2$) in the final matrix $B$.

Linear algebra must provide a way to see what happened. The result of elimination was $T = A^TA = LDL^T$, and the matrix above with orthogonal columns is $B = A(L^T)^{-1}$. First we verify that $B^TB$ is diagonal: if $A^TA = LDL^T$ then $B^TB = L^{-1}A^TA(L^T)^{-1} = D$. The lengths squared of the columns, on the diagonal of $B^TB$, are the pivots. The row operations on $T$ were a premultiplication by $L^{-1}$, and the column operations on $A$ are postmultiplication by $(L^T)^{-1}$. The final step is to *normalize the columns*, dividing by the square roots of 2, 3/2, 4/3, to reach $Q = A(L^T)^{-1}D^{1/2}$. Its connection to $A$ will be clearer if we get rid of the inverses: *Gram-Schmidt factors $A$ into the product of $Q$ (with orthonormal columns) and $R = D^{1/2}L^T$ (upper triangular with positive diagonal).*

The triangularity of $R$ reflects the special order in which the process operates: it subtracts multiples of earlier columns from later columns, and never changes a column after it is orthogonalized. That is like elimination, which subtracts earlier rows from later rows; a row is never changed after its leading entries have been annihilated. The triangular $L$ connects $U$ to $T$, and the triangular $R$ connects $Q$ to $A$.

### 3. The inverses of $L$ and $T$

The inverse matrices that were avoided above are not so bad. They will be full of nonzeros, so they are not right for computations—but they add to the understanding. To invert $L$ we write it as $I - N$; the strictly triangular $N$ contains $1/2$ and $2/3$. The inverse of $I - N$ is $I + N + N^2 + \cdots$, which is a finite sum because $N^3 = 0$:

$$^{-1} = \begin{bmatrix} 1 & & \\ \frac{1}{2} & 1 & \\ \frac{1}{3} & \frac{2}{3} & 1 \end{bmatrix}.$$

In a larger matrix the pattern is more striking. This makes an excellent example in teaching, since $N$ contains $1/2, 2/3, 3/4, \cdots$ and $N^2$ contains $1/3, 2/4, \cdots$ and $N^3$ contains $1/4, 2/5, \ldots$. *The nonzero entries of* $L^{-1}$ *are* $j/i$. I should have seen that coming, but it was unexpectedly neat.

We could invert $T$ directly from its cofactors (which are computable), or from the product $(L^T)^{-1}D^{-1}L^{-1}$. There is also an indirect approach, which brings out the parallels with differential equations. The matrix $T$ multiplies the $j$th column of $T^{-1}$ to give the $j$th column of the identity matrix. In the continuous case, minus the second derivative operates on the Green's function to give a delta function. *The analogue of* $TT^{-1} = I$ *is* $(-d^2/dx^2)G = \delta$. That is a textbook example, solvable because $G$ is linear away from the spike in the delta function:

$$G = \begin{matrix} x(1-y) & \text{for } x \leqslant y \\ y(1-x) & \text{for } y \leqslant x \end{matrix} \qquad T_{ij}^{-1} = \begin{matrix} \dfrac{i}{n+1}\left(1 - \dfrac{j}{n+1}\right) & \text{for } i \leqslant j \\ \dfrac{j}{n+1}\left(1 - \dfrac{i}{n+1}\right) & \text{for } j \leqslant i \end{matrix}.$$

The first derivative of $G$ drops by one at $x = y$, so the second derivative is $-\delta$. In the matrix case, the first difference down each column of $T^{-1}$ drops by one at the diagonal $i = j$. That becomes visible when we include the zero boundary conditions at $i = 0$ and $i = 4$:

$$\begin{array}{c} 0 \quad 0 \quad 0 \\ T^{-1} = \dfrac{1}{4}\begin{bmatrix} 3 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix}. \\ 0 \quad 0 \quad 0 \end{array}$$

The determinant $n + 1 = 4$ appears in the denominator, and this explicit form establishes a further property—that $T^{-1}$ is a *positive matrix*. If the forces $f$ are all downward, so are the displacements $x = T^{-1}f$.

## 4. Eigenvalues, eigenvectors, and the singular value decomposition

It is the eigenvectors of $T$ that are most interesting, and they imitate the continuous case. There the eigenfunctions of $-d^2/dx^2$ are $\sin j\pi x$, vanishing at the endpoints. Here they are *discrete sines*. Their components $\sin ij\pi/(n+1)$ vanish at $i = 0$ and $i = n+1$. After normalization they are the columns of the eigenvector matrix $S$:

$$S = \frac{1}{\sqrt{2}}\begin{bmatrix} \sin\theta & \sin 2\theta & \sin 3\theta \\ \sin 2\theta & \sin 4\theta & \sin 6\theta \\ \sin 3\theta & \sin 6\theta & \sin 9\theta \end{bmatrix} \quad \text{with} \quad \theta = \pi/(n+1) = \pi/4.$$

This "sine matrix" is symmetric and orthogonal: $S^T = S^{-1} = S$.

The eigenvalues of $T$ are $2 \pm \sqrt{2}$ and 2. In general they are $\lambda_j = 2 - 2\cos j\theta$, in the interval from 0 to 4. They appear when $T$ multiplies the eigenvectors, and in principle they give another approach to the solution of $Tx = f$—expand $f$ as a combination of eigenvectors $s_j$ and divide by the eigenvalues: $x = \Sigma(f, s_j)s_j/\lambda_j$ or $x = S\Lambda^{-1}S^{-1}f$. Normally that is a crazy choice compared to elimination. Here elimination is quick because $T$ is tridiagonal, but we mention that the Fast Fourier Transform multiplies by $S$ and $S^{-1}$ in $n \log n$ steps. (In a two-dimensional problem the FFT defeats elimination [6].) The worst of all would be to multiply by $T^{-1}$! It is full of nonzeros and $T^{-1}f$ needs $n^2$ separate multiplications.

The eigenvectors of $A^TA$, and the square roots of the eigenvalues, are two of the three factors in the *singular value decomposition of* $A$. That factorization is important [1], [5] and becoming better known. Like $LU$ and $QR$, the decomposition $A = C\Sigma S^T$ is the matrix analogue of a key factorization in the theory of Lie algebras—where the decompositions are associated with Bruhat, Iwasawa, and Cartan [2]. The other factor $C$ is the eigenvector matrix for $AA^T$:

$$AA^T = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}.$$

This corresponds to four springs and three masses, with no supports at the ends. Its extra eigenvalue is $\lambda = 0$, with $(1,1,1,1)$ as eigenvector. The other eigenvectors are also even functions—discrete cosines rather than discrete sines, in the columns of the orthogonal matrix

$$C = \frac{1}{\sqrt{2}}\begin{bmatrix} \cos\theta/2 & \cos\theta & \cos 3\theta/2 & 1/\sqrt{2} \\ \cos 3\theta/2 & \cos 3\theta & \cos 9\theta/2 & 1/\sqrt{2} \\ \cos 5\theta/2 & \cos 5\theta & \cos 15\theta/2 & 1/\sqrt{2} \\ \cos 7\theta/2 & \cos 7\theta & \cos 21\theta/2 & 1/\sqrt{2} \end{bmatrix}, \qquad \theta = \frac{\pi}{n+1} = \frac{\pi}{4}.$$

Now we put together the singular value decomposition, apologizing for brevity (there is a second family to come!). Multiplying $A = C\Sigma S^T$ by its transpose gives

$$A^TA = S\Sigma^T\Sigma S^T \quad \text{and} \quad AA^T = C\Sigma\Sigma^TC^T.$$

The 4 by 3 matrix $\Sigma$ is diagonal, with entries $(2-\sqrt{2})^{1/2}$, $2^{1/2}$, and $(2+\sqrt{2})^{1/2}$. Then $\Sigma^T\Sigma$ contains the three eigenvalues of $A^TA$, $\Sigma\Sigma^T$ contains the four eigenvalues of $AA^T$ (including zero), and all we have done is to diagonalize those two matrices by $S$ and $C$.

## 5. Sums of squares

Returning just for a moment to the positive definite $T$, it should be possible to express $x^TTx = 2x_1^2 - 2x_1x_2 + 2x_2^2 - 2x_2x_3 + 2x_3^2$ as a sum of squares. One way is to take care of the cross product terms and add what is needed:

$$x^TTx = (x_2 - x_1)^2 + (x_3 - x_2)^2 + x_1^2 + x_3^2.$$

A more systematic way is to complete squares:

$$x^TTx = 2\left(x_1 - \frac{1}{2}x_2\right)^2 + \frac{3}{2}\left(x_2 - \frac{2}{3}x_3\right)^2 + \frac{4}{3}x_3^2.$$

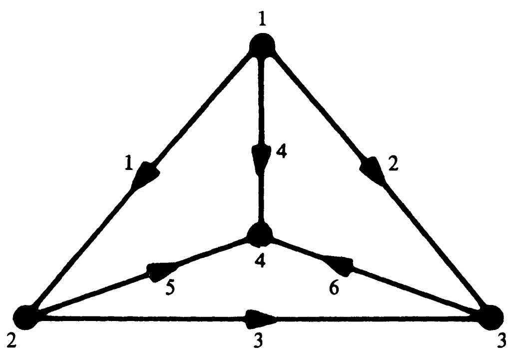
A third way must come from the eigenvectors:

$$x^TTx = \sum_1^3 \lambda_j (x_1 \sin j\theta + x_2 \sin 2j\theta + x_3 \sin 3j\theta)^2/2.$$

Those correspond exactly to $A^TA$ and $LDL^T$ and $S\Lambda S^T$. The first one uses four squares, because $A$ has four rows. It separates $x^TTx$ into contributions from the four springs, while $LDL^T$ gives Lagrange's sum of squares and $S\Lambda S^T$ is the spectral theorem. Always the entries in the rows are inside the squares, and the diagonals of $I$ and $D$ and $\Lambda$ are outside.

### The Second Family of Matrices

We come back to applied mathematics. The line of springs connected only nearest neighbors. Suppose we include *all* edges between nodes, to create a network of pipes or resistors. There are $N = 4$ nodes and $(1/2)N(N-1) = 6$ edges, so the incidence matrix $A_0$ is 6 by 4:




$$\begin{array}{rc}
 & \begin{array}{cccc} & & & \end{array} \quad \text{edge} \\
A_0 = & \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \\
\text{node} & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array}
\end{array}$$

Each row of $A_0$ corresponds to an edge of the graph. The entries $\pm 1$ are in columns that correspond to nodes on the edge ($-1$ for the starting node, $+1$ for the ending node). The construction describes the "connectivity" of the network, and we display one member of the second family:

$$K_0 = A_0^TA_0 = \begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{bmatrix}.$$

This matrix is *singular*. Its nullspace contains $(1, 1, 1, 1)$, because that vector is in the nullspace of $A_0$. The columns of $A_0$ add to the zero column, and the equation

$e = A_0x$ has no unique solution. It is impossible to determine the four $x$'s from a knowledge of their six differences (the $e$'s). At least one component of $x$ must be fixed by a boundary condition, and we set $x_4 = 0$. In an electrical network, node 4 has been grounded. In a network of pipes, node 4 is a reservoir. The potential or the pressure is zero, and *the last column is removed from* $A_0$. The resulting 6 by 3 matrix, with independent columns, will be called $A$.

The incidence matrix $A$ controls flow in the network. It enters in both of Kirchhoff's fundamental laws:

*voltage law*: the potential differences around a loop add to zero
*current law*: the currents into a node add to zero.

For pipes, the pressure differences around every loop add to zero—and the flow into a node equals the flow out. The first law makes the potential or the pressure well defined, and it connects $x$ to $e$. The second law is the continuity equation, or "conservation law," and it connects $y$ to $f$. This is the same pattern $e = Ax$ and $f = A^Ty$ that we saw for springs and masses, and it is repeated throughout applied mathematics [6]. *The basic form of equilibrium rests on* $A^TCA$.

The equation $e = Ax$ is clear. The action of $A$, with entries $-1$ and $+1$, is to compute differences in pressure or potential—which go into $e$. The dual equation is $f = A^Ty$, and it looks down the columns of $A$ (or along the rows of $A^T$). Each column corresponds to a node, and the plus or minus signs measure current in or out. The currents $y$ are balanced by the external sources $f$, and conservation at the three nodes is

$$\begin{aligned} f_1 &= -y_1 - y_2 - y_4 \\ f_2 &= \phantom{-}y_1 - y_3 - y_5 \\ f_3 &= \phantom{-}y_2 + y_3 - y_6 \end{aligned} \quad \text{or} \quad f = \begin{bmatrix} -1 & -1 & 0 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 & 0 & -1 \end{bmatrix} y.$$

*The matrix is* $A^T$. It is interesting to see the continuous analogues: the voltage law says there is no rotation (curl $e = 0$) and the current law says there is conservation ($-\operatorname{div} y = f$). Then $A$ completes the great triad from vector calculus: *it is the gradient*, and $e$ is the gradient of the potential. (Gradients have zero curl, and $(\text{gradient})^T = -\text{divergence}$.) The divergence theorem integrates $-\operatorname{div} y = f$ to find the flow out through the boundary. The discrete divergence theorem adds the $f$'s to find the flow out through the ground.†

As before, we need a third law $y = Ce$. It expresses the physical properties of the six edges, through a diagonal matrix $C$. The entries $c_1c_2 \cdots c_6$ come from Ohm's law (for resistors) and Poiseuille's law (for pipes). As in Hooke's law or Newton's law, the first and most basic approximation is linearity—but the pattern does not require it, and Einstein showed that it is not quite correct. For real conductors or real pipes or real springs $y = Ce$ is nearly true for small $e$ (as it is in relativity), and the framework is

$$\underset{x}{\text{potentials}} \xrightarrow[A]{} \underset{e}{\text{differences}} \xrightarrow[C]{} \underset{y}{\text{currents}} \xrightarrow[A^T]{} \underset{f}{\text{source}} \quad \text{or} \quad A^TCAx = f.$$

†Adding the three equations above gives $f_1 + f_2 + f_3 = -y_4 - y_5 - y_6$.

Finally we compute the symmetric positive definite matrix

$$K = A^TCA = \begin{bmatrix} c_1+c_2+c_4 & -c_1 & -c_2 \\ -c_1 & c_1+c_3+c_5 & -c_3 \\ -c_2 & -c_3 & c_2+c_3+c_6 \end{bmatrix}.$$

The diagonal entries indicate which edges touch the nodes (including edges to ground). If all conductances are $c_i = 1$ then the entries are 3 and $-1$, and $K$ is a submatrix of the earlier $K_0 = A_0^TA_0$—which included row 4 and column 4 and was singular. Fixing $x_4 = 0$ made the matrix $K$ positive definite.

Our second family contains the singular $K_0$ ($N$ by $N$) and the nonsingular $K$ ($n$ by $n$), always from a complete graph with $C = I$ and $n = N - 1$:

$$K_0 = \begin{bmatrix} N-1 & -1 & \cdot & -1 \\ -1 & N-1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & -1 \\ -1 & \cdot & -1 & N-1 \end{bmatrix}$$

and

$$K = \begin{bmatrix} n & -1 & \cdot & -1 \\ -1 & n & \cdot & \cdot \\ \cdot & \cdot & \cdot & -1 \\ -1 & \cdot & -1 & n \end{bmatrix}.$$

Those are $A_0^TA_0$ and $A^TA$. We do the computations for $N = 4$ and $n = 3$.

## 6. Elimination and determinants and spanning trees

Elimination on $K$ produces a multiple of the next smaller $K$:

$$\begin{bmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{bmatrix} \to \begin{bmatrix} 3 & -1 & -1 \\ & 8/3 & -4/3 \\ & -4/3 & 8/3 \end{bmatrix}.$$

The multipliers at the first stage are $l_{i1} = -1/n$, later $l_{ij} = -1/(n+1-j)$:

$$K = \begin{bmatrix} 1 & & \\ -\frac{1}{3} & 1 & \\ -\frac{1}{3} & -\frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} 3 & & \\ & \frac{8}{3} & \\ & & 2 \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{3} & -\frac{1}{3} \\ & 1 & -\frac{1}{2} \\ & & 1 \end{bmatrix} = LDL^T.$$

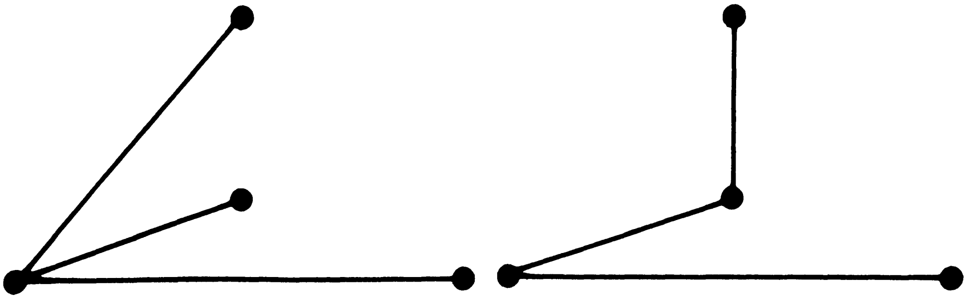
The pivots on the diagonal are not hard to track down; for the $n$ by $n$ matrix they are

$$\frac{n}{n+1}, \frac{n-1}{n}, \cdots \frac{2}{3}, \frac{1}{2}, \quad \text{all multiplied by } n+1.$$

Their product is $(n+1)^{n-1}$, which is the determinant of $K_n$. That is a celebrated formula in graph theory (Kirchhoff's formula), which *counts the number of spanning trees* in the network:

*A spanning tree connects all nodes to ground, without loops.*

There are $4^2 = 16$ of them for the complete graph with four nodes:



4 trees are "stars from a node" $(1/2)4! = 12$ trees are "paths"

The reason for this determinant formula is in the incidence matrix $A$. Its square submatrices $S$ (3 by 3) have determinant $+1$ or $-1$ when the rows of $S$ correspond to edges of a spanning tree. If the edges contain loops, $S$ is singular and $\det S = 0$. Then Laplace's expansion of the determinant of a product produces a count of the spanning trees:

$$\det A^TA = \sum_{\text{submatrices } S} (\det S^T)(\det S) = \sum_{\text{spanning trees}} 1$$

That is a little outside linear algebra, and I was puzzled when thinking back to the first family. The springs were in a line, and *where are the spanning trees in a straight line graph*? If any edge along the line is skipped, the graph becomes disconnected—and spanning trees are usually connected. Nevertheless there must be $n + 1$ spanning trees, equal to the determinant of $T_n$.

The key is that *two* nodes were grounded. The line of $n + 1$ springs was fixed at both ends. Therefore deleting any spring leaves a spanning tree—all nodes are connected to one ground or the other. Note that the whole line, which connects ground to ground, counts as a loop! Only the $n + 1$ broken lines are spanning trees, and that number is the correct determinant.

If the line is fixed at one end, the whole graph becomes a spanning tree and the others are gone. In that case $\det A^TA = 1$. Our description of spanning trees had to differ from the usual one, because there were multiple grounds. If the grounded nodes are *identified*, the definition becomes standard again.

## 7. The inverses of $K$ and $L$

$K$ is an easy matrix to invert. It is $(n + 1)I - ee^T$, a multiple of the identity perturbed by a rank one matrix. The vector $e^T = [1 \cdots 1]$ gives the off-diagonal $-1$'s. For such a matrix the inverse has the same form: $K^{-1}$ is $I/(n + 1) + cee^T$ for some number $c$. Multiplying $K$ by $K^{-1}$ leads to the inner product $e^Te = n$ and to the right value for $c$:

$$I + \left[(n+1)c - \frac{1}{n+1} - nc\right]ee^T = I \quad \text{or} \quad c = \frac{1}{n+1}.$$

Thus the inverse matrix is $K^{-1} = (I + ee^T)/(n + 1)$, again positive:

$$K^{-1} = \begin{bmatrix} n & -1 & \cdot & -1 \\ -1 & n & \cdot & -1 \\ \cdot & \cdot & \cdot & -1 \\ -1 & -1 & -1 & n \end{bmatrix}^{-1} = \frac{1}{n+1}\begin{bmatrix} 2 & 1 & \cdot & 1 \\ 1 & 2 & \cdot & 1 \\ \cdot & \cdot & \cdot & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix}.$$

The electrical interpretation is straightforward: $x = K^{-1}f$ gives the potentials at the nodes. Suppose there is a current into the first node, emerging at the grounded node: $f = (1, 0, \dots, 0)$. By symmetry all other nodes will be at the same potential, halfway between node 1 and the zero potential at the ground. That matches $x = K^{-1}f$, which is the first column of $K^{-1}$ and is proportional to $(2, 1, \dots, 1)$.

The lower triangular $L$ is more interesting to invert. I don't know why you would do it, but the result is amazing. Its pattern is visible for $n = 3$ but more beautiful for $n = 5$:

$$L = \begin{bmatrix} 1 & & & & \\ -\frac{1}{5} & 1 & & & \\ -\frac{1}{5} & -\frac{1}{4} & 1 & & \\ -\frac{1}{5} & -\frac{1}{4} & -\frac{1}{3} & 1 & \\ -\frac{1}{5} & -\frac{1}{4} & -\frac{1}{3} & -\frac{1}{2} & 1 \end{bmatrix} \quad \text{and} \quad L^{-1} = \begin{bmatrix} 1 & & & & \\ \frac{1}{5} & 1 & & & \\ \frac{1}{4} & \frac{1}{4} & 1 & & \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 1 & \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 1 \end{bmatrix}.$$

It is not even clear why $LL^{-1} = I$! Looking at row 5 of $L$ and column 1 of $L^{-1}$, there is a telescoping sum:

$$\begin{aligned} &-\frac{1}{5} - \frac{1}{4\cdot 5} - \frac{1}{3\cdot 4} - \frac{1}{2\cdot 3} + \frac{1}{2} \\ &\quad = -\frac{1}{5} + \left(\frac{1}{5} - \frac{1}{4}\right) + \left(\frac{1}{4} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{2}\right) + \frac{1}{2} = 0. \end{aligned}$$

Other inner products vanish in the same way. Or we could multiply in the order $L^{-1}L$, and prove by induction that the product is the identity. There must be an elegant way to see that whole construction.

*Note:* $L^{-1}$ comes from elimination on $K^{-1}$, but not in the usual order. The factorization $K^{-1} = (L^{-1})^T D^{-1} L^{-1}$ eliminates from the *bottom row upward*, and it reaches a *lower triangular* matrix—which is $L^{-1}$ after dividing out the pivots. Looking back at $K^{-1}$, the last row of $L^{-1}$ is immediately confirmed as $\frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2}, 1$.

## 8. The eigenvalues of $K$ and $K_0$

The eigenvalues of the $n$ by $n$ matrix of 1's (which is $ee^T$) are $0, \dots, 0$ and $n$. That is typical of a rank one matrix: there must be $n - 1$ zero eigenvalues and the last eigenvalue equals the trace. The eigenvector for $\lambda = n$ is $e$ itself, and the other eigenvectors are orthogonal to $e$ (because the matrix is symmetric).

For $K = (n-1)I - ee^T$ the eigenvalues are shifted—they are $n + 1, \dots, n + 1$ and $n + 1 - n = 1$. Their product is $(n+1)^{n-1}$, equal to the determinant! For $K_0 = NI - ee^T$ the eigenvalues are $N, \dots, N$ and $N - N = 0$. It is singular as expected.

The condition number $\lambda_{\max}(K)/\lambda_{\min}(K)$, which controls the sensitivity of $Kx = f$, is $n + 1$. That is well below the condition number of $T$, which grows like

$n^2$. If $T$ corresponds to $-d^2/dx^2$, then $Kx = f$ is closer to the integral equation $x(t) - \int_0^1 x(s)\,ds = f(t)$ with $x(1) = 0$.

By a fluke we already have an orthogonal matrix $Q$ that diagonalizes $K$ or $K_0$. Its last column contains the eigenvector $e$ (normalized to unit length). Its other columns must contain vectors that are perpendicular to $e$ *and to each other*. That is exactly the construction that was carried out for the first family of matrices. In the line of springs, the columns of $A$ were $\ldots, 1, -1, \ldots$ (perpendicular to $(1, 1, \ldots, 1)$) and those columns were orthogonalized. Therefore the result is an orthogonal set of eigenvectors for $K$ or $K_0$:

$$Q = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & 1 \\ -1 & \frac{1}{2} & \frac{1}{3} & 1 \\ 0 & -1 & \frac{1}{3} & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{2} & & & \\ & \sqrt{3/2} & & \\ & & \sqrt{4/3} & \\ & & & \sqrt{4} \end{bmatrix}^{-1}.$$

## 9. The nullspace of $A^T$

It was easy to find the nullspace of $A_0$. It contained the vector $e = (1, \ldots, 1)$ of equal potentials, for which the potential differences are all zero: $A_0 e = 0$. Its nullspace disappeared when the last node was grounded and the last column of $A_0$ was removed, leaving $A$. Now we want the solutions to $A_0^T y = 0$ or to $A^T y = 0$, which have the same solutions since a dependent row has no effect. We are looking for currents $y_1, \ldots, y_6$ that satisfy Kirchhoff's law, with flow in equal to flow out and no sources.

The answer is in the loops of the graph. *If current flows around a loop, Kirchhoff's conservation law is satisfied.* The 4-node graph drawn earlier has three small loops, and watching the arrows as we go around those triangles gives the loop currents

$$y_1 = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}, \quad y_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \\ 0 \\ 1 \end{bmatrix}, \quad y_3 = \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \\ 1 \\ -1 \end{bmatrix}.$$

The large loop around the outside triangle is a combination of these inner loops, and adds nothing new. Since the rank of $A_0$ and $A$ is $n = N - 1$, and the number of rows (edges in the graph) is $m$, the nullspaces of $A^T$ and $A_0^T$ must have dimension $m - (N - 1)$. That leads to a linear algebra proof of Euler's formula: *number of independent loops* $= m - N + 1$. If several nodes are grounded then linear algebra still gives the correct number of loops, properly redefined. (They connect a grounded node to a grounded node, without self-intersections.) That number is the dimension of the nullspace of $A^T$—equal to 1 for the line of springs.

To close we look again at those vectors $y_1$, $y_2$, $y_3$ from the loops. They appear to be the columns of a new incidence matrix! They are, and the fourth column $(1, -1, 1, 0, 0, 0)$ completes it with $+1$ and $-1$ in every row. It is the *incidence matrix $A_D$ for the dual graph*, and the extra column comes from the outside triangle —the loop around infinity that is not independent of the others. The loops of the dual graph are the nodes of the original graph, and linear algebra is taking us into a

new pattern—with a matrix $A_D^T C^{-1} A_D$ that must somehow be dual to $A^T CA$. I hesitate to go further.

*Note added in proof.* I did not attempt to orthogonalize the columns of $T$, since there was no special reason for hope. But André Weideman's computer kept the faith. Down each orthogonalized column, it showed entries which ***increase linearly*** up to the final nonzero. For $n = 4$, the columns are proportional to those of $M$:

$$T = \begin{bmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 2 \end{bmatrix} \to M = \begin{bmatrix} 1 & 1 & 1 & 1 \\ a & 2 & 2 & 2 \\ & b & 3 & 3 \\ & & c & 4 \end{bmatrix}.$$

Why is $M$ the right matrix? Its $j$th column is a combination of columns $1, \dots, j$ of $T$. Furthermore, ***column $j$ of $M$ is orthogonal to columns*** $1, \dots, j-1$ ***of*** $T$. $T$ is a second-difference matrix, the second differences of a linear function are zero!

**Acknowledgement.** This research was supported by the National Science Foundation (DMS 84-03222) and the Army Research Office (DAAG29-83-K0025 and DAAL03-86-K0171).

REFERENCES

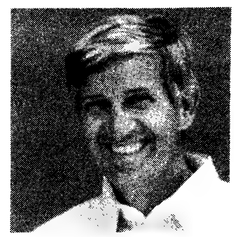
1. G. Golub and C. Van Loan, Matrix Computations, Johns Hopkins Press, Baltimore, 1983.
2. R. Howe, Very basic Lie theory, *Amer. Math. Monthly*, 90 (1983) 600–623.
3. R. T. Rockafellar, Network Flows and Monotropic Optimization, John Wiley, New York, 1984.
4. G. Strang, A framework for equilibrium equations, *SIAM Review*, 30 (1988) 283–97.
5. ______, Linear Algebra and Its Applications, Academic Press, 1980, 3rd edition, Harcourt Brace Jovanovich, 1988.
6. ______, Introduction to Applied Mathematics, Wellesley-Cambridge Press, Box 157, Wellesley, MA, 1986.

# Chebyshev's Inequality and Natural Density

CURTIS N. COOPER AND ROBERT E. KENNEDY

*Curtis N. Cooper* has been at Central Missouri State University since 1978, where he is a Professor of Mathematics and Computer Science. He received his Ph.D. from Iowa State University in 1978. His main interests lie in number theory and numerical analysis.

*Robert E. Kennedy* has been at Central Missouri State University since 1967, where he is a Professor of Mathematics. He received his Ph.D. from the University of Missouri at Kansas City in 1973. His main interests lie in number theory and commutative algebra.



**1. Introduction and Notation.** Sometimes in resolving a particular open question, the method used can be generalized in order to investigate other questions. For example, in [1] the concept of a "Niven number" was introduced and investigated. (A Niven number is a positive integer which is divisible by its digital sum.) In [2], it was shown that the natural density of the set of Niven numbers is 0. In what follows, we generalize the method used in [2] to give sufficient conditions in order that certain sets of integers, namely integers $n$ divisible by a function $f(n)$, have a natural density zero. First, we give the following notation and definitions.

Let $f$ be an integer valued function. For an integer $x$, define the following sets of nonnegative integers:

$$S = \{n : f(n) \text{ divides } n\},$$
$$[0, x) = \{n : 0 \leqslant n < x\},$$

and the integer

$$S(x) = \#(S \cap [0, x)).$$

Let the sequence $f(0), f(1), f(2), \ldots, f(x-1)$ be denoted by $f([0, x))$. Note that $f([0, x))$ is a random variable taking on the values of the above sequence. That is, we are interested not only in $f(k)$, but also in the frequency of $f(k)$. Then, letting $\mu$ and $\sigma^2$ be the mean and variance of $f([0, x))$, respectively, we may write

$$\mu = \frac{1}{x} \sum_{0 \leqslant k < x} f(k)$$

and

$$\sigma^2 = \frac{1}{x} \sum_{0 \leqslant k < x} (f(k))^2 - \mu^2.$$

For $k > 0$, we let

$$\begin{aligned} I_k &= \{ n \in f([0, x)) : |n - \mu| < k\sigma \}, \\ A &= \{ n \in [0, x) : n \text{ is a multiple of a member of } I_k \}, \\ B &= \{ n \in [0, x) : |f(n) - \mu| \geqslant k\sigma \}. \end{aligned}$$

**2. An Upper Bound for $S(x)/x$.** Here, we recall Chebyshev's Inequality [**3**; Chapter 8]:

Let $X$ be a random variable with mean $\mu$ and variance $\sigma^2$. Then for each $k > 0$,

$$\text{Prob}[|X - \mu| \geqslant k\sigma] \leqslant \frac{1}{k^2}.$$

Stating this theorem in terms of the notation given above, we see, that for any $k > 0$,

$$\#B \leqslant \frac{x}{k^2}.$$

It also follows by the description of $A$ that

$$\#A \leqslant \sum_{t \in I_k} \left[ \frac{x}{t} \right],$$

where as usual, the square brackets denote the integral part operator. We will henceforth restrict $k$ so that $k < \mu/\sigma$ in order that $\mu - k\sigma > 0$.

Referring to the definitions of $S$, $A$, and $B$, we see that

$$S \cap [0, x) \subseteq A \cup B$$

for any positive integer $x$. This follows since if $n$ is an element of $S \cap [0, x)$, then $f(n)$ is a factor of $n$ less than $x$. If $f(n)$ is an element of $I_k$, then $n$ is an element of $A$, while on the other hand, $f(n)$ not an element of $I_k$ implies that $n$ is an element of $B$. So, in either case, $n$ is an element of $A \cup B$.

Thus, it is immediate that

$$S(x) \leqslant \#A + \#B.$$

Therefore, for each $k$ such that $0 < k < \mu/\sigma$, we have

$$S(x) \leqslant \sum_{t \in I_k} \left[ \frac{x}{t} \right] + \frac{x}{k^2}. \tag{2.1}$$

Continuing, we remove the square brackets and integrate to obtain

$$S(x) \leqslant \int_{\mu - k\sigma}^{\mu + k\sigma} \left( \frac{x}{t} \right) dt + \frac{x}{\mu - k\sigma} + \frac{x}{k^2}. \tag{2.2}$$

The second term on the right side of (2.2) is necessary since the summands of (2.1) are strictly decreasing and hence, the first term of (2.1) is not taken into considera-

tion when the sum is replaced by an integral. Hence, we now have

$$\frac{S(x)}{x} \leqslant \ln\left(\frac{\mu + k\sigma}{\mu - k\sigma}\right) + \frac{1}{\mu - k\sigma} + \frac{1}{k^2}, \tag{2.3}$$

which gives an upper bound for $S(x)/x$.

**3. The Natural Density of $S$.** Using (2.1), we can prove the following theorem which gives sufficient conditions in order that

$$\lim_{x \to \infty} \frac{S(x)}{x} = 0,$$

where $x$ ranges over the integers. That is, sufficient conditions in order that the natural density of $S$ is zero will be given.

THEOREM 1. *Let $f$, $S$, $\mu$, and $\sigma$ be as defined above. If*

$$\lim_{x \to \infty} \frac{\mu}{\sigma} = \infty$$

*and*

$$\lim_{x \to \infty} \mu = \infty,$$

*then*

$$\lim_{x \to \infty} \frac{S(x)}{x} = 0.$$

*Proof.* (outline) The theorem follows by letting $k = (\mu/\sigma)^{1/2}$ in (2.3) and taking the limit of each side. Actually, we may choose the exponent to be any positive number less than one.

If $a_n$ is the $n$th term of an increasing sequence, a more general, and sometimes more useful, theorem can be similiarly proven. Its proof is based upon observing that for $a_n \leqslant x < a_{n+1}$, we have

$$\frac{S(x)}{x} \leqslant \frac{S(a_{n+1})}{a_n} \cdot \frac{a_{n+1}}{a_{n+1}} = \frac{S(a_{n+1})}{a_{n+1}} \cdot \frac{a_{n+1}}{a_n}.$$

THEOREM 2. *Let $a_n$ be the nth term of an increasing sequence and for each $n$, let $\mu$ = mean $f([0, a_n))$ and $\sigma^2$ = variance $f([0, a_n))$ for an integer valued function $f$. Then for $S$ and $S(x)$ as defined above,*

$$\lim_{n \to \infty} \mu = \infty,$$

$$\lim_{n \to \infty} \frac{\mu}{\sigma} = \infty,$$

*and if $a_{n+1}/a_n$ is bounded, then*

$$\lim_{x \to \infty} \frac{S(x)}{x} = 0.$$

**4. The Natural Density of the Niven Numbers.** Hence, sufficient conditions have been given for sets such as $S$ to have a natural density of zero. Here, an outline of the proof that the natural density of the Niven numbers is zero will be given. This

was proven in [2] but the outline given here makes use of the "big-O" notation which, in our opinion makes for a more elegant proof.

We let $s(n)$ be the digital sum of $n$ and for an integer $x$, consider the set $[0, x)$. Then the mean $\mu$, and variance $\sigma^2$, of $s([0, x))$ are given by

$$\mu = 4.5 \log x + 0(1) \tag{4.1}$$

and

$$\sigma^2 = 0(\log x), \tag{4.2}$$

respectively, where $\log x$ denotes the common logarithm. The proof of these statistics is found in [4] and [5].

So, using (4.1), (4.2) and Theorem 1, we have the following theorem.

THEOREM 3. *Let $N(x)$ be the number of Niven numbers not exceeding $x$. Then the natural density of $N$ is zero, that is*

$$\lim_{x\to\infty} \frac{N(x)}{x} = 0.$$

*Proof.* By (4.1) and (4.2), we have that

$$\lim_{x\to\infty} \mu = \infty$$

and

$$\lim_{x\to\infty} \frac{\mu^2}{\sigma^2} = \lim_{x\to\infty} \frac{20.25 \log^2 x + 0(\log x)}{0(\log x)} = \infty.$$

Therefore,

$$\lim_{x\to\infty} \frac{\mu}{\sigma} = \infty,$$

and by Theorem 1, the natural density of the Niven numbers is zero.

**5. Little $\omega$ numbers, log numbers and square root numbers.** We observe that Theorems 1 and 2 can be used to study the natural density of other sets of numbers. Consider the following integer valued functions of the set of nonnegative integers.

$\omega(n)$ = number of distinct prime factors of $n$ where $\omega(1)$ is defined to be 1,
$l(n) = [\log_b n]$, for $n \geqslant 1$ and

$$r(n) = [n^{1/2}].$$

Here, we will investigate the sets

$$W = \{n : \omega(n) \text{ divides } n\} = \{1, 2, 3, 4, \ldots, 14, 16, \ldots\},$$

$$L = \{n : l(n) \text{ divides } n\} = \{b, b+1, b+2, \ldots, b^2 - 1, \ldots\}$$

and

$$R = \{n : r(n) \text{ divides } n\} = \{1, 2, 3, 4, 6, 8, \ldots\}$$

with respect to the question of natural density. In what follows, $W$, $L$, and $R$ are called the set of "little $\omega$ numbers," "log numbers," and "square root numbers," respectively.

For any $x$, it can be shown [**6**, pp. 355–357] that

$$\sum_{n \leqslant x} \omega(n) = x \ln(\ln x) + 0(x)$$

and

$$\sum_{n \leqslant x} (\omega(n))^2 = x(\ln(\ln x))^2 + 0(x \ln(\ln x)).$$

Hence, it follows that the mean, $\mu$, and variance, $\sigma^2$, of $\omega([0, x))$ are $\ln(\ln x) + 0(1)$ and $0(\ln(\ln x))$ respectively. Thus,

$$\lim_{x \to \infty} \mu = \infty$$

and since

$$\lim_{x \to \infty} \frac{\mu^2}{\sigma^2} = \infty,$$

we also have that

$$\lim_{x \to \infty} \frac{\mu}{\sigma} = \infty.$$

Therefore, by Theorem 1, the natural density of $W$ is 0.

In considering the set, $L$, of log numbers base $b$, we investigate the interval $[1, b^n]$ for any positive integer $n$. Then

$$\begin{aligned}
\mu = \text{mean of } l([1, b^n]) &= \frac{1}{b^n} \sum_{1 \leqslant x \leqslant b^n} l(x) \\
&= \frac{1}{b^n}\left(\sum_{t=0}^{n-1} \sum_{x=b^t}^{b^{t+1}-1} l(x) + n\right) \\
&= \frac{1}{b^n}\left(\sum_{t=0}^{n-1} \sum_{x=b^t}^{b^{t+1}-1} t + n\right) \\
&= \frac{1}{b^n}\left((b-1) \sum_{t=0}^{n-1} t b^t + n\right).
\end{aligned}$$

But,

$$(b-1) \sum_{t=0}^{n-1} t b^t = \frac{(n-1)b^{n+1} - nb^n + b}{b-1}.$$

Therefore,

$$\mu = n\left(1 + \frac{1}{b^n}\right) + \frac{1}{b^{n-1}(b-1)} - \frac{b}{b-1},$$

which for convenience, we will write as $\mu = n + 0(1)$. Similarly it can be shown that

$$\begin{aligned}
\frac{1}{b^n} \sum_{x=1}^{b^n} (l(x))^2 = n^2\left(1 + \frac{1}{b^n}\right) - \frac{2bn}{b-1} + \frac{b(b+1)}{(b-1)^2} \\
- \frac{1}{b^{n-2}(b-1)^2} - \frac{1}{b^{n-1}(b-1)^2} = n^2 + 0(n).
\end{aligned}$$

Therefore,

$$\begin{aligned}\sigma^2 &= \text{variance of } l([1, b^n]) \\ &= n^2 + 0(n) - (n + 0(1))^2 = 0(n).\end{aligned}$$

Since

$$\lim_{n\to\infty} \mu = \infty,$$

$$\lim_{n\to\infty} \frac{\mu}{\sigma} = \infty,$$

and $b^{n+1}/b^n = b$ is bounded, we have by Theorem 2 that the natural density of $L$ is zero.

Finally, to demonstrate that the conditions given by Theorem 1 and Theorem 2 are not necessary, we consider the square root numbers. Let $n$ be a positive integer. Then

$$\begin{aligned}\mu = \text{mean } r\big([0, n^2)\big) &= \frac{1}{n^2} \sum_{x=0}^{n^2-1} r(x) \\ &= \frac{1}{n^2} \sum_{t=0}^{n-1} \sum_{x=t^2}^{(t+1)^2-1} r(x) \\ &= \frac{1}{n^2} \sum_{t=0}^{n-1} t(2t+1) \\ &= \frac{2n}{3} - \frac{1}{2} - \frac{1}{6n}\end{aligned}$$

which we write as $2n/3 + 0(1)$. By a similar technique, it can be shown that

$$\frac{1}{n^2} \sum_{x=0}^{n^2-1} (r(x))^2 = \left(\frac{1}{2}\right)n^2 + 0(n).$$

Therefore

$$\begin{aligned}\sigma^2 = \text{variance } r\big([0, n^2)\big) &= \frac{1}{n^2} \sum_{x=0}^{n^2-1} (r(x))^2 - \mu^2 \\ &= \frac{n^2}{18} + 0(n).\end{aligned}$$

Noting that

$$\lim_{n\to\infty} \mu = \infty,$$

but that

$$\lim_{n\to\infty} \frac{\mu}{\sigma} = 2\sqrt{2},$$

we see that Theorem 1 cannot be applied in the investigation of the square root numbers with respect to natural density. However, that the natural density of $R$ is zero follows from [7, Solution to E 2491, pp. 854–855] which notes that the $n$th

square root number is given by the formula

$$a_n = \left(\left[\frac{n-1}{3}\right] + 1\right)\left(n - 2\left[\frac{n-1}{3}\right]\right),$$

from which it follows that the natural density of $R$ is zero.

**6. Conclusion.** It is possible to investigate other integer valued functions with respect to natural density by using Theorems 1 and 2. In particular, another set which we believe could be investigated is what could be called the set of "big omega numbers." To describe this set, let

$$n = p_1^{n_1} p_2^{n_2} \cdots p_m^{n_m}$$

be the canonical representation of the positive integer $n$ and define the integer valued function

$$\Omega(n) = n_1 + n_2 + n_3 + \cdots + n_m.$$

Thus, we call an integer $n$ a big omega number if $\Omega(n)$ divides $n$. It is known [**6**; pp. 355–357] that for any $x$

$$\text{mean}\,\Omega([1, x)) = \ln(\ln x) + 0(1).$$

However, we have not determined the variance of $\Omega([1, x))$ at this time. So, we cannot make use of Theorem 1 or Theorem 2 yet. We leave the question of the natural density of the set of big omega numbers as an open question.

The authors thank the referee for the many helpful and valuable comments, corrections, and suggestions.

REFERENCES

1. R. Kennedy, T. Goodman, and C. Best, Mathematical discovery and Niven numbers, the *MATYC Journal*, 14 (1980) 21–25.
2. R. Kennedy and C. Cooper, On the natural density of the Niven numbers, the *College Mathematics Journal*, 15 (1984) 309–312.
3. S. Ross, A First Course in Probability, 2nd ed., Macmillan, New York, 1984.
4. P. Cheo and S. Yien, A problem on the $K$-adic representation of positive integers, *Acta Math. Sinica*, 5 (1955) 433–438.
5. R. Kennedy and C. Cooper, The Variance of the Digital Sum Function, CMSU preprint (1985).
6. G. Hardy and E. Wright, An Introduction to the Theory of Numbers, 4th ed., Oxford Press, New York, 1960.
7. E. Dixon and R. Mueller (independently), Problem Solution, this MONTHLY, 82 (1975) 854–855.

# The Editor's Corner: The New Mersenne Conjecture

P. T. BATEMAN, J. L. SELFRIDGE*, AND S. S. WAGSTAFF, JR.**

It is well known that Mersenne stated in his *Cogitata* [4] that, of the fifty-five primes $p \leqslant 257$, $2^p - 1$ is itself prime only for the eleven values

$$p = 2, 3, 5, 7, 13, 17, 19, 31, 67, 127, \text{ and } 257.$$

It is also well known that his list had five errors: $p = 67$ and $257$ should have been removed from the list while $p = 61$, $89$, and $107$ should have been added to it.

Several authors [**1**, **2**, **3**] have speculated about how Mersenne formed his list. It is easy to notice that all numbers on his (incorrect) list lie within 3 of some power of 2. However, Mersenne certainly knew that $2^{11} - 1$ is composite and hence that not all primes $p = 2^k \pm 3$ produce prime $M_p = 2^p - 1$. The next prime of this form not on Mersenne's list is $p = 29$. He surely knew that $M_{29}$ is composite, as it has the small divisor 233. Also 263 divides $2^{131} - 1$. Mersenne's list is explained by the rule

$$M_p \text{ is prime if and only if } p \text{ is a prime of one of the forms } 2^k \pm 1 \text{ or } 2^{2k} \pm 3 \quad (1)$$

except for the omission of $p = 61$. In fact Mersenne stated in [**5**, Chap. 21, p. 182] a rule very similar to (1). (The verb "differs"—not "exceeds," as some have guessed—is omitted from his sentence, but Mersenne supplied it in a corrigendum on the back of page 235.) Drake [**2**] quotes this sentence from [**5**], locates the missing verb and argues that (1) was in fact Mersenne's rule. He suggests that 61 was missing from [**4**] either because of a typographical error or because Mersenne mistakenly believed that $M_{61}$ is composite. When copying a list, like "$\ldots, 61, 67, \ldots$", containing two adjacent similar items, it is a common error to omit the first of these (here "61").

Now the question presents itself: Is there a neat way to distinguish the Mersenne hits like 31, 61, 127 from the Mersenne misses like $67, 257, \ldots$ and $89, 107, \ldots$? When $(2^{127} + 1)/3$ was proved prime, we began looking at the other $(2^p + 1)/3$. We noticed that they were prime for the hits and composite for the misses! Is this accidental? Will "a little more computing" find a counterexample?

We replace (1) by this new, related conjecture that when both sides of (1) are true, $(2^p + 1)/3$ is prime, and when (1) is false, $(2^p + 1)/3$ is composite. Restating this conjecture we get the

NEW MERSENNE CONJECTURE. *If two of the following statements about an odd positive integer p are true, then the third one is also true.*

(a) $p = 2^k \pm 1$ or $p = 4^k \pm 3$.
(b) $M_p$ is prime.
(c) $(2^p + 1)/3$ is prime.

It is not necessary to assume that $p$ is prime, for if $p$ is composite (or 1), then statements (b) and (c) are both false and the conjecture holds.

It is easy to find examples of primes $p$ for which all three statements are true ($p = 3, 5, 7, 13, 17, 19, 31, 61, 127$) or all three are false ($p = 29, 37, 41, 47, \ldots$) or

*Department of Mathematical Sciences, Northern Illinois University, DeKalb, IL 60115
**Department of Computer Sciences, Purdue University, West Lafayette, IN 47907

Table for "The New Mersenne Conjecture"

| $p$ | $p = 2^k \pm 1$ or $4^k \pm 3$? | $2^p - 1$ prime? | $(2^p + 1)/3$ prime? |
|---|---|---|---|
| 3 | yes (−1) | yes | yes |
| 5 | yes (+1) | yes | yes |
| 7 | yes (−1 or +3) | yes | yes |
| 11 | no | no: 23 | yes |
| 13 | yes (−3) | yes | yes |
| 17 | yes (+1) | yes | yes |
| 19 | yes (+3) | yes | yes |
| 23 | no | no: 47 | yes |
| 31 | yes (−1) | yes | yes |
| 43 | no | no: 431 | yes |
| 61 | yes (−3) | yes | yes |
| 67 | yes (+3) | no: 193707721 | no: 7327657 |
| 79 | no | no: 2867 | yes |
| 89 | no | yes | no: 179 |
| 101 | no | no: 7432339208719 | yes |
| 107 | no | yes | no: 643 |
| 127 | yes (−1) | yes | yes |
| 167 | no | no: 2349023 | yes |
| 191 | no | no: 383 | yes |
| 199 | no | no: 164504919713 | yes |
| 257 | yes (+1) | no: 535006138814359 | no: 37239639534523 |
| 313 | no | no: 10960009 | yes |
| 347 | no | no: 14143189112952632419639 | yes |
| 521 | no | yes | no: 510203 |
| 607 | no | yes | no: 115331 |
| 701 | no | no: 796337 | yes |
| 1021 | yes (−3) | no: 40841 | no: 10211 |
| 1279 | no | yes | no: 706009 |
| 1709 | no | no: 379399 | yes |
| 2203 | no | yes | no: 13219 |
| 2281 | no | yes | no: 22811 |
| 2617 | no | no: 78511 | yes |
| 3217 | no | yes | no: 7489177 |
| 3539 | no | no: 7079 | yes (prp) |
| 4093 | yes (−3) | no | no |
| 4099 | yes (+3) | no: 73783 | no: 2164273 |
| 4253 | no | yes | no: 118071787 |
| 4423 | no | yes | no |
| 8191 | yes (−1) | no: 338193759479 | no |
| 9689 | no | yes | no: 19379 |
| 9941 | no | yes | no |
| 11213 | no | yes | no |
| 16381 | yes (−3) | no | no: 163811 |
| 19937 | no | yes | no |
| 21701 | no | yes | no: 43403 |
| 23209 | no | yes | no: 4688219 |
| 44497 | no | yes | no: 2135857 |
| 65537 | yes (+1) | no | no |
| 65539 | yes (+3) | no | no: 58599599603 |
| 86243 | no | yes | no |
| 110503 | no | yes | no |
| 131071 | yes (−1) | no: 231733529 | no: 2883563 |
| 132049 | no | yes | no |
| 216091 | no | yes | no |
| 262147 | yes (+3) | no: 268179002471 | no: 4194353 |
| 524287 | yes (−1) | no: 62914441 | no |

exactly one is true ($p = 67, 257, 1021, \ldots$ for only (a) true; $p = 89, 107, 521, \ldots$ for only (b) true; and $p = 11, 23, 43, 79, \ldots$ for only (c) true). However, the New Mersenne Conjecture is true for all $p$ less than 100000, which is the current limit of the search for Mersenne primes. It is valid also for all $p$ between $10^5$ and $10^6$ for which at least one of the three statements is known to hold. We expect that the three statements are true simultaneously only for the nine primes mentioned above.

The Table above summarizes what is known about our conjecture. It lists all odd primes $p$ satisfying at least one of these three conditions:

(1) $p < 1000000$ and $p = 2^k \pm 1$ or $p = 4^k \pm 3$.
(2) $p < 100000$ and $2^p - 1$ is prime.
(3) $p < 4000$ and $(2^p + 1)/3$ is prime.

When a number is asserted to be composite, a factor is given if one is known. The factors of $M_{131071}$ and $M_{524287}$ were found by Robinson [6]. The 1065-digit number $(2^{3539} + 1)/3$ passed a probabilistic primality test, but we did not give a complete proof that it is prime.

It is a simple consequence of quadratic reciprocity that if $p \equiv 1 \pmod 4$, then the factors of $2^p - 1$ are congruent to 1 or $6p + 1 \pmod{8p}$, and if $p \equiv 3 \pmod 4$, then the factors of $2^p - 1$ are congruent to 1 or $2p + 1 \pmod{8p}$. This observation is the starting point for a heuristic argument [7] which concludes that the number of $p$ less than $y$ for which $M_p$ is prime is about $e^\gamma \log_2 y \approx 1.78 \log_2 y$, where $\gamma$ is Euler's constant.

Likewise, one can show that if $p \equiv 1 \pmod 4$, then the factors of $(2^p + 1)/3$ are congruent to 1 or $2p + 1 \pmod{8p}$, and if $p \equiv 3 \pmod 4$, then the factors of $(2^p + 1)/3$ are congruent to 1 or $6p + 1 \pmod{8p}$. A heuristic argument like the one mentioned above concludes that the number of $p$ less than $y$ for which $(2^p + 1)/3$ is prime is also about $e^\gamma \log_2 y$.

The total number of natural numbers less than $y$ with one of the forms $2^k \pm 1$ or $4^k \pm 3$ is about $3 \log_2 y$. Hence, the number of primes less than $y$ with one of these forms is $O(\log y)$.

In view of the foregoing heuristics and the fact that there are about $y/\log y$ primes less than $y$, the probability that any one of the three statements holds for a randomly chosen prime $p$ less than $y$ is $O(y^{-1} \log^2 y)$. If the three statements were independent random events, then the expected number of primes $p$ greater than $L$ for which at least two of the statements hold is about $C\int_L^\infty y^{-2} \log^4 y\,dy$, which is finite. Substituting $L = 100000$ gives an upper bound on the expected number of failures of the New Mersenne Conjecture. Assuming a reasonable value for $C$ (about 9) we find that the expected number of failures is less than 1. This is one reason why we believe that the conjecture is true. Another reason is that it holds for all $p$ less than 100000 as well as those larger $p$ for which it has been tested.

We are grateful to Duncan A. Buell and Jeff Young for testing the primality of $(2^p + 1)/3$ for several $p > 50000$, using a Cray 2 computer.

REFERENCES

1. R. C. Archibald, Mersenne's numbers, *Scripta Math.*, 3 (1935), 113.
2. Stillman Drake, The rule behind 'Mersenne's numbers', *Physis–Riv. Internaz. Storia Sci.*, 13 (1971) 421–424. MR 58#26870.

3. Malcolm R. Heyworth, A conjecture on Mersenne's conjecture, *New Zealand Math. Mag.*, 19 (1982) 147–151. MR 85a:11002.
4. M. Mersenne, Cogitata Physico Mathematica, Parisiis, 1644, Praefatio Generalis No. 19.
5. ______, Novarum Observationum Physico-Mathematicarum, Tomus III, Parisiis, 1647.
6. Raphael M. Robinson, Some factorizations of numbers of the form $2^n \pm 1$, *Math. Tables Aids Comput.* 11 (1957) 265–268, MR 20 #832.
7. S. S. Wagstaff, Jr., Divisors of Mersenne numbers, *Math. Comp.*, 40 (1983) 385–397, MR 84j: 10052.

# UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

*In this department the* MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

## Polynomials All of Whose Derivatives Have Integer Roots

C. E. CARROLL
*Department of Physics, Auburn University, Alabama 36849*

Let $\mathscr{P}_n$ be the set of $n$th degree polynomials $p(x)$ such that $p, p', p'', \ldots, p^{(n-1)}$ all have integer roots. The superscripts stand for differentiation. $\mathscr{P}_n$ is not the empty set, because all the roots of $p(x)$ can be the same integer. Assuming that $n > 1$, can we find other members of $\mathscr{P}_n$? This question was recently raised by Frank Schmidt [5], but there is evidence of its being older. Kloke, Pethö, Carroll, and others each noted that

$$x^{n-1}(x - n) \tag{1}$$

is a polynomial in $\mathscr{P}_n$ [3]. Furthermore, if $m$ is any integer, then any nonzero multiple of

$$q_n(x) = x^{n-1}(x + mn) \tag{2}$$

is in $\mathscr{P}_n$. This is proved by computing

$$q_n'(x) = nq_{n-1}(x).$$

If the conjecture given below is correct, we can find all the members of $\mathscr{P}_n$ and answer Schmidt's latest question [6].

Three obvious properties of $\mathscr{P}_n$ are (i) $p(x) \in \mathscr{P}_n \Rightarrow \alpha p(x) \in \mathscr{P}_n$ for real $\alpha \neq 0$, (ii) $p(x) \in \mathscr{P}_n \Rightarrow p(x - a) \in \mathscr{P}_n$ for integer $a$, and (iii) $p(x) \in \mathscr{P}_n \Rightarrow p' \in \mathscr{P}_{n-1}$. We can show that $p(x) \in \mathscr{P}_n$ implies that the average of all roots of $p(x)$ is an integer. Let $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots$ have roots $\lambda_1, \ldots, \lambda_n$. Then $\Sigma\lambda_i$ is one of the elementary symmetric polynomials [4] and $a_{n-1} = -a_n(\Sigma\lambda_i)$. Repeated use of (iii) gives $p^{(n-1)} = a_n(nx - \Sigma\lambda_i)(n-1)! \in \mathscr{P}_1$, which implies that $\Sigma\lambda_i/n$ is an integer. Property (ii) implies that $p(x + \Sigma\lambda_i/n) \in \mathscr{P}_n$. Thus, any polynomial in $\mathscr{P}_n$ may be reduced to a polynomial of the form $a_n x^n + a_{n-2}x^{n-2} + \cdots$. When finding all members of $\mathscr{P}_n$, we may assume $\Sigma\lambda_i = 0$, without loss of generality. Polynomials (1) and (2) have $x = 0$ as the multiple root, unless $n < 3$, but we may not assume that each member of $\mathscr{P}_n$ has a multiple root.

If $n < 3$, the members of $\mathscr{P}_n$ are easy to find. $\mathscr{P}_3$ was treated several years ago by Bruggeman and Gush [2], and it appears in the recent problem proposed by P. A. Batnik [1]. Here, we are concerned mainly with $\mathscr{P}_4$. If a member of $\mathscr{P}_4$ has two double roots, we use $\Sigma\lambda_i = 0$ and property (i) to write

$$p(x) = (x + m)^2(x - m)^2, \tag{3}$$

where $m$ is a nonzero integer. The roots of $p''$ are $\pm 3^{-1/2}\, m$, which cannot be integers. This contradiction shows that no member of $\mathscr{P}_4$ has two double roots. If $p(x) \in \mathscr{P}_4$ and has fewer than three distinct roots, then $p(x)$ has a triple or quadruple root. We drop the condition $\Sigma\lambda_i = 0$, in order to put the multiple root at $x = 0$. Thus, $p(x)$ is proportional to

$$x^3(x + 4m), \tag{4}$$

where $m$ is any integer; this is a special case of (2). The successive derivatives of (4) are $4x^2\,(x + 3m)$, $12x\,(x + 2m)$, and $24\,(x + m)$.

CONJECTURE. *If $p(x) \in \mathscr{P}_4$ and has more than two distinct roots, then it is proportional to*

$$(x - 193\,m)(x - 141\,m)(x + 167\,m)^2, \tag{5}$$

*where $m$ is a nonzero integer. The assumption that $\Sigma\lambda_i = 0$ has been used.*

Evidence for this conjecture is provided by two computer programs. One program examines all sets of four integers $\lambda_1, \ldots, \lambda_4$ such that $\Sigma\lambda_i = 0$ and $\Sigma\lambda_i^2 <$ 4993532. The other program does a more extensive and efficient search, by finding pairs of integer roots of polynomials $p(x)$ such that $p' \in \mathscr{P}_3$. The polynomial (5) appears in some unpublished work by M. S. Klamkin, around 1950, but he did not consider higher-degree polynomials.

If this conjecture is correct, we can find all members of $\mathscr{P}_n$, where $n > 4$. Each of these polynomials has a root of multiplicity $n - 1$ or $n$. If we put the multiple root at $x = 0$, every such polynomial is a multiple of (2). To prove this, we start with $\mathscr{P}_5$. If $p(x) \in \mathscr{P}_5$ and has only single and double roots, then $p' \in \mathscr{P}_4$ and $p'$ has four distinct roots, which is contrary to the conjecture. If $p(x) \in \mathscr{P}_5$ and has a triple root, then $p'$ has a double root; assuming $\Sigma\lambda_i = 0$, we find that $p'$ is a multiple of (5). Hence, $p(x)$ is a multiple of

$$(x + 167\,m)^3(x - k)(x - l),$$

where $k$, $l$, $m$ are integers, and $m \neq 0$. Some computation gives

$$(k - l)^2 = -41975\,m^2.$$

This shows that $p(x)$ cannot have a triple root. If $p(x) \in \mathscr{P}_5$ and the conjecture is correct, then $p(x)$ has a quadruple or quintuple root. Putting the multiple root at $x = 0$, we find that $p'$ is proportional to (4) and $p(x)$ is proportional to $q_5(x)$, given by (2). Similar reasoning makes any member of $\mathscr{P}_6$ into a multiple of $q_6(x)$. The argument is easily extended to higher degrees.

REFERENCES

1. P. A. Batnik, problem E3221, this MONTHLY, 94 (1987) 681.
2. T. Bruggeman and T. Gush, Nice cubic polynomials for curve sketching, *Math. Mag.*, 53 (1980) 233–234.
3. Kloke, Pethö and Carroll, solutions of 86-5, *Math. Intelligencer*, 9 No. 3 (1987) 43.
4. S. Lang, Algebra, Addison-Wesley, Reading, Massachusetts, 1971, p. 132.
5. F. Schmidt, problem 86-5, *Math. Intelligencer*, 8 No. 2 (1986) 48.
6. ______, problem 87-9, *Math. Intelligencer*, 9 No. 3 (1987) 40.

# NOTES

EDITED BY DAVID J. HALLENBECK, DENNIS DETURCK, AND ANITA E. SOLOW

## Generalizing the Formula for Areas of Polygons to Moments

S. F. BOCKMAN
*Department of Aeronautics and Astronautics, Stanford University, Stanford, CA 94305*

**1. Introduction.** There is a well-known formula for the area of an arbitrary polygon whose vertices are given in Cartesian coordinates. For example, it is available in the CRC Standard Mathematical Tables [1]. This formula has been found useful in surveying at least since the availability of desktop calculating machines [2]. Readers of this journal were reminded of the formula's existence [3]. The purpose of this note is to point out the motivation for, and the ease of, generalizing the formula to give moments of arbitrary polygons. I do not know of any reference where such formulae appear.

Often when the area of a polygon is needed higher moments are needed as well. For a dynamics problem not only the area, but the centroid and moments of inertia, of a thin plate may be needed. In order to compute the bending behavior of a beam the centroid and moments of inertia of its cross section are needed. Simple formulae giving the moments of a polygon in terms of boundary points are particularly useful in computer programming. They allow convenient approximation of the moments of an arbitrary plane area. Since the formulae are correct for nonconvex areas, and can be trivially generalized to non-simply connected areas, they reduce the need for complex program logic.

My desire for moment formulae analagous to the area formula arose a few years ago from a problem in biomechanics. The dynamics properties of a structure in the inner ear of the Greater Horse Shoe Bat were to be computed. The area, centroid, and moment of inertia of a nonconvex cross section were required. A large number of points on the boundary of the cross section could conveniently be digitized from an illustration in a journal article. The area of the resulting polygon was immediately available from the well-known area formula. But, how were the centroid and polar moment of inertia to be found? Clearly some sort of brute force method would work. However, the elegant approach would be to use formulae like that for the area. Unfortunately, the references at hand did not contain such formulae.

After some attempts to generalize the area formula to higher moments it became clear that the effort had the flavor of a proof of the planar case of Stokes's Theorem (cf. [3]). In fact, generalizations could be easily obtained by applying Green's Theorem to the polygon assuming appropriate functions.

**2. Moment formulae.** Recall the planar case of Stokes's Theorem. For a simply connected area in the plane and functions $P(x, y)$ and $Q(x, y)$ the following equality holds

$$\int_{\text{area}} (-dP/dy + dQ/dx)\, dx\, dy = \int_{\text{boundary}} (P\, dx + Q\, dy).$$

The area formula can be reproduced by applying the theorem to the polygon and

making the choice, $P = 0$, $Q = x$. Then

$$\int_{\text{area}} da = \int_{\text{area}} (-dP/dy + dQ/dx)\, dx\, dy = \int_{\text{boundary}} x\, dy.$$

If the polygon's vertices are given in $x$-$y$ coordinates and numbered counterclockwise from 1 to $n$, then doing the required integration yields,

$$\text{Area} = (1/2) \sum_{i=2}^{N+1} (x_i + x_{i-1})(y_i - y_{i-1}),$$

where we define $x_{N+1} = x_1$ and $y_{N+1} = y_1$. Note that a perhaps more common form of the formula can be reproduced by dropping canceling terms of the form $x_i y_i$,

$$\text{Area} = (1/2)(x_1 y_2 - x_2 y_1 + x_2 y_3 - x_3 y_2 + \cdots + x_N y_1 - x_1 y_N).$$

Similarly, the first and second moments with respect to the $x$-axis, and the mixed-term second moment, can be obtained with the choice $P = 0$, and the choices $Q = x^2/2$, $Q = x^3/3$, and $Q = x^2 y/2$, respectively. The results of the integrations are

$$\int x\, dx\, dy = (1/6) \sum_{i=2}^{N+1} \left(x_i^2 + x_i x_{i-1} + x_{i-1}^2\right)(y_i - y_{i-1})$$

$$\int x^2\, dx\, dy = (1/12) \sum_{i=2}^{N+1} \left(x_i^3 + x_i^2 x_{i-1} + x_i x_{i-1}^2 + x_{i-1}^3\right)(y_i - y_{i-1})$$

$$\int xy\, dx\, dy = (1/24) \sum_{i=2}^{N+1} \left\{\left[y_i\left(3x_i^2 + 2x_i x_{i-1} + x_{i-1}^2\right)\right.\right.$$

$$\left.\left. + y_{i-1}\left(x_i^2 + 2x_i x_{i-1} + 3x_{i-1}^2\right)\right](y_i - y_{i-1})\right\}.$$

The rest of the first and second moments can be easily obtained from those given above. Clearly the method easily generalizes to higher moments. It should be noted that the particular summands obtained are not unique since the choices of $P$ and $Q$ are not unique.

REFERENCES

1. William H. Beyer (ed.), CRC Standard Mathematical Tables, 27th Edition, CRC Press, Boca Raton, FL, 1984. p. 194.
2. Harry Bouchard and Francis H. Moffit, Surveying, 5th edition, International Textbook Company, Scranton, PA, 1965, p. 228.
3. M. G. Stone, A mnenomic for areas of polygons, *Amer. Math. Monthly*, 93 (1986) 479–480.

# An Elementary Test for the Galois Group of a Quartic Polynomial

LUISE-CHARLOTTE KAPPE
*Department of Mathmatical Sciences, SUNY at Binghamton, Binghamton, NY 13901*

BETTE WARREN
*Department of Mathematics, Eastern Michigan University, Ypsilanti, MI 48197*

**1. Introduction.** The problem of determining the Galois group of a polynomial from its coefficients has held the interest of mathematicians for over a hundred years. There is a classical algorithm for determining the Galois group of a polynomial from its roots which can be found in [4] as well as [8] and [11]. But as van der Waerden notes in [11], the method is cumbersome and is not of much interest from a practical point of view. More recently Richard Stauduhar [10] has applied modern insights to old techniques to develop and implement a computer algorithm that finds Galois groups of low degree polynomials with integer coefficients.

We are concerned with the case of quartic polynomials. The earliest complete account which we have found is contained in the unpublished dissertation of 1895 by F. Hack [3], as cited in [4]. Other analyses of the problem were published during the first quarter of this century ([1], [2], and [9]). More modern treatments are given by I. Kaplansky [7, p. 52], T. Hungerford [5, p. 273], and N. Jacobson [6, p. 253].

The test which we present in section 2 can be used to determine the Galois group of any irreducible quartic polynomial over any field $\mathbf{K}$ of characteristic other than 2. It is based on the same reduction to the determination of the Galois group of the cubic resolvent polynomial used in [7] and [5]. However, we avoid the need to check irreducibility of the quartic over extensions of the base field in order to distinguish between the cases of the cyclic and dihedral Galois groups by using a modification suggested by Frank Stephanic. Our computational requirement is the ability to solve quadratic equations and factor cubic polynomials over the base field $\mathbf{K}$, both of which are elementary procedures in the common case where the base field is assumed to be the rational numbers, $\mathbf{Q}$.

In section 3 we consider the special case of even quartic polynomials $\mathbf{p}(x) = x^4 + bx^2 + d$, which, following [8], we call "biquadratic." (However, the reader should be warned that the term "biquadratic" is used as a synonym for "quartic" in many other sources.) We give an irreducibility criterion for such polynomials (Theorem 2), and the test of section 2 is used to formulate a simple mechanical procedure for determining the Galois group of an irreducible biquadratic polynomial directly either from its coefficients or from its easily obtainable roots (Theorem 3). Similar results can be found in [7] and [8]. In the last section the simplicity and usefulness of the tests is illustrated by examples and by giving construction methods for some biquadratic polynomials with all possible Galois groups.

**2. The Elementary Test.** Let $\mathbf{p}(x) = \prod_{1 \leqslant i \leqslant 4}(x - \alpha_i) = x^4 + ax^3 + bx^2 + cx + d$ be a polynomial with splitting field $\mathbf{F}$ over a base field $\mathbf{K}$ and let $\mathbf{K}^2 = \{k^2 | k \in \mathbf{K}\}$ denote the set of all squares of elements in $\mathbf{K}$. If $\mathbf{p}(x)$ is irreducible and separable (no repeated roots), then its Galois group, Gal($\mathbf{F}/\mathbf{K}$), acts transitively on the four roots and must therefore be isomorphic to one of the transitive subgroups

of $\mathbf{S}_4$: a Klein 4-group, $\mathbf{V}$, a cyclic group of order 4, $\mathbf{C}_4$, a dihedral group of order 8, $\mathbf{D}_4$, the alternating group, $\mathbf{A}_4$, or $\mathbf{S}_4$ itself. We assume throughout this note that the characteristic of $\mathbf{K}$ is not 2. This guarantees that $\mathbf{p}(x)$ will be separable if it is irreducible.

The polynomial $\mathbf{r}(x) = x^3 - bx^2 + (ac - 4d)x - (a^2d - 4bd + c^2)$ whose roots are $t_1 = \alpha_1\alpha_2 + \alpha_3\alpha_4$, $t_2 = \alpha_1\alpha_3 + \alpha_2\alpha_4$ and $t_3 = \alpha_1\alpha_4 + \alpha_2\alpha_3$ is called the cubic resolvent of $\mathbf{p}(x)$. Its splitting field in $\mathbf{F}$ will be denoted by $\mathbf{E}$. Finally, we note that $\mathbf{r}(x)$ and $\mathbf{p}(x)$ have the same discriminant, $D = \prod_{i<j}(\alpha_i - \alpha_j)^2 = \prod_{i<j}(t_i - t_j)^2$ or $D = -4b^3A + b^2B^2 + 18dBA - B^3 - 27A^2$ with $A = a^2d + c^2 - 4bd$ and $B = ac - 4d$ (see [**6**, p. 251]).

THEOREM 1. *Let* $\mathbf{K}$ *be a field of characteristic* $\neq 2$. *Suppose* $\mathbf{p}(x)$ *is irreducible over* $\mathbf{K}$, $\mathbf{r}(x)$ *its cubic resolvent with splitting field* $\mathbf{E}$, *and* $D$ *the discriminant. Then*

(i) $\mathrm{Gal}(\mathbf{F}/\mathbf{K}) \cong \mathbf{S}_4$ *if and only if* $\mathbf{r}(x)$ *is irreducible over* $\mathbf{K}$ *and* $D \notin \mathbf{K}^2$;

(ii) $\mathrm{Gal}(\mathbf{F}/\mathbf{K}) \cong \mathbf{A}_4$ *if and only if* $\mathbf{r}(x)$ *is irreducible over* $\mathbf{K}$ *and* $D \in \mathbf{K}^2$;

(iii) $\mathrm{Gal}(\mathbf{F}/\mathbf{K}) \cong \mathbf{V}$ *if and only if* $\mathbf{r}(x)$ *splits into linear factors over* $\mathbf{K}$;

(iv) $\mathrm{Gal}(\mathbf{F}/\mathbf{K}) \cong \mathbf{C}_4$ *if and only if* $\mathbf{r}(x)$ *has exactly one root* $t$ *in* $\mathbf{K}$ *and* $\mathbf{g}(x) = (x^2 - tx + d)(x^2 + ax + (b - t))$ *splits over* $\mathbf{E}$;

(v) $\mathrm{Gal}(\mathbf{F}/\mathbf{K}) \cong \mathbf{D}_4$ *if and only if* $\mathbf{r}(x)$ *has exactly one root* $t$ *in* $\mathbf{K}$ *and* $\mathbf{g}(x)$ *does not split over* $\mathbf{E}$.

*Proof.* Let $\mathbf{V}_4 = \{(1), (12)(34), (13)(24), (14)(23)\}$, the only transitive subgroup of $\mathbf{S}_4$ which is isomorphic to $\mathbf{V}$. The splitting field $\mathbf{E}$ of the cubic resolvent $\mathbf{r}(x)$ is contained in $\mathbf{F}$. Since $\mathbf{p}(x)$ is separable and irreducible, the common discriminant $D$ is not zero, and $\mathbf{r}(x)$ has distinct roots, regardless of whether or not $\mathbf{r}(x)$ is irreducible (see [**5**, p. 325]). It can be verified by direct calculation that a permutation of $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ fixes all three roots of $\mathbf{r}(x)$ if and only if it acts on the subscripts of the $\alpha$'s as an element of $\mathbf{V}_4$. Thus $\mathrm{Gal}(\mathbf{F}/\mathbf{E}) = \mathrm{Gal}(\mathbf{F}/\mathbf{K}) \cap \mathbf{V}_4$ and we see that

1) $\mathbf{r}(x)$ splits into linear factors over $\mathbf{K}$ if and only if $\mathrm{Gal}(\mathbf{F}/\mathbf{K}) \subset \mathbf{V}_4$;

2) $\mathbf{r}(x)$ is irreducible over $\mathbf{K}$ if and only if $|\mathrm{Gal}(\mathbf{F}/\mathbf{K})| = |\mathrm{Gal}(\mathbf{F}/\mathbf{E})|\,|\mathrm{Gal}(\mathbf{E}/\mathbf{K})|$ is divisible by 3.

Since $\mathbf{p}(x)$ is irreducible, $|\mathrm{Gal}(\mathbf{F}/\mathbf{K})|$ is divisible by 4, hence its order is at least 4. Thus in the first case we have $\mathrm{Gal}(\mathbf{F}/\mathbf{K}) \cong \mathbf{V}$, while in the second case $|\mathrm{Gal}(\mathbf{F}/\mathbf{K})|$ is divisible by 12, so $\mathrm{Gal}(\mathbf{F}/\mathbf{K}) \cong \mathbf{A}_4$ or $\mathbf{S}_4$. In particular, $\mathrm{Gal}(\mathbf{F}/\mathbf{K})$ will be $\mathbf{A}_4$ if the discriminant $D = \prod_{i<j}(\alpha_i - \alpha_j)^2$ is a square in $\mathbf{K}$ and $\mathbf{S}_4$ if it is not (since $\mathbf{K}$ does not have characteristic two and $D \neq 0$, $\sqrt{D} \neq -\sqrt{D}$).

Now let us assume that $\mathbf{r}(x)$ has exactly one root, $t$, in $\mathbf{K}$. After possibly renumbering we may assume that $t = t_1 = \alpha_1\alpha_2 + \alpha_3\alpha_4$. In light of the previous paragraph we have $\mathrm{Gal}(\mathbf{F}/\mathbf{K}) \cong \mathbf{C}_4$ or $\mathbf{D}_4$. Consider the polynomial $\mathbf{g}(x) = (x^2 - tx + d)(x^2 + ax + (b - t))$ over $\mathbf{K}$. The roots of its first factor are $\alpha_1\alpha_2$ and $\alpha_3\alpha_4$, and the roots of the second are $\alpha_1 + \alpha_2$ and $\alpha_3 + \alpha_4$. If $\mathrm{Gal}(\mathbf{F}/\mathbf{K}) \cong \mathbf{C}_4$ then $\mathbf{E}$ is the only quadratic extension of $\mathbf{K}$ contained in $\mathbf{F}$. Thus each of the quadratic factors of $\mathbf{g}(x)$ splits in $\mathbf{E}$. Conversely, assume that $\mathbf{g}(x)$ splits completely over $\mathbf{E}$. Then the polynomial $\mathbf{k}(x) = x^2 - (\alpha_1 + \alpha_2)x + \alpha_1\alpha_2$ with roots $\alpha_1, \alpha_2$ has coefficients in $\mathbf{E}$. Let $\mathbf{M}$ be the splitting field of $\mathbf{k}(x)$ over $\mathbf{E}$. Now $\mathbf{E} \subseteq \mathbf{M} \subseteq \mathbf{F}$, and hence $\alpha_1$, $\alpha_2$, $t_2$, $t_3$ and $t_1 = t$ are in $\mathbf{M}$. Since $\alpha_3 + \alpha_4 = -a - (\alpha_1 + \alpha_2)$ it follows that $\alpha_3 + \alpha_4 \in \mathbf{M}$. Furthermore, $\alpha_1 - \alpha_2 \neq 0$ together with $(\alpha_1 - \alpha_2)(\alpha_3 - \alpha_4) = t_2 - t_3$ imply that $\alpha_3 - \alpha_4$ is in $\mathbf{M}$. Hence $\alpha_3, \alpha_4 \in \mathbf{M}$, since $\operatorname{char}\mathbf{K} \neq 2$. This yields

$\mathbf{F} = \mathbf{M}$ and hence $|\mathrm{Gal}(\mathbf{F}/\mathbf{M})| = 1$. Since

$$|\mathbf{C}_4| \leqslant |\mathrm{Gal}(\mathbf{F}/\mathbf{K})| = |\mathrm{Gal}(\mathbf{F}/\mathbf{M})|\,|\mathrm{Gal}(\mathbf{M}/\mathbf{E})|\,|\mathrm{Gal}(\mathbf{E}/\mathbf{K})| \leqslant 1 \cdot 2 \cdot 2 = 4 < |\mathbf{D}_4|,$$

$\mathrm{Gal}(\mathbf{F}/\mathbf{K})$ must be isomorphic to $\mathbf{C}_4$.

The condition for $\mathbf{D}_4$ follows by mutual exclusion. □

We suggest that the reader pause now and determine the Galois groups of the following polynomials over $\mathbf{Q}$: $\mathbf{p}_1(x) = x^4 - 4x^3 + 4x^2 + 6$, $\mathbf{p}_2(x) = x^4 + 4x^3 + 6x^2 + 4x + 2$, $\mathbf{p}_3(x) = x^4 + x^3 + x^2 + x + 1$ (see Example 1), and $\mathbf{f}(x) = x^4 + px + p$ for each prime $p$ (see Example 2).

**3. An Application to Biquadratic Polynomials.** In this section we apply our test to irreducible polynomials of the form $\mathbf{h}(x) = x^4 + bx^2 + d$ with $b, d \in \mathbf{K}$, a field of characteristic $\neq 2$. First we prove the following irreducibility criterion for such polynomials.

THEOREM 2. *Let $\mathbf{h}(x) = x^4 + bx^2 + d$ be a polynomial over $\mathbf{K}$, a field of characteristic $\neq 2$, and $\pm\alpha, \pm\beta$ its roots. Then the following conditions are equivalent*:

(i) *$\mathbf{h}(x)$ is irreducible over $\mathbf{K}$*;

(ii) $\alpha^2, \alpha + \beta, \alpha - \beta \notin \mathbf{K}$;

(iii) $b^2 - 4d, -b + 2\sqrt{d}, -b - 2\sqrt{d} \notin \mathbf{K}^2$.

*Proof.* The polynomial $h(x)$ cannot have an irreducible cubic factor, so it is reducible over $\mathbf{K}$ if and only if at least one of the factorizations

$$\mathbf{h}(x) = (x^2 - \alpha^2)(x^2 - \beta^2),$$

$$\mathbf{h}(x) = \big(x^2 - (\alpha + \beta)x + \alpha\beta\big)\big(x^2 + (\alpha + \beta)x + \alpha\beta\big),$$

$$\mathbf{h}(x) = \big(x^2 - (\alpha - \beta)x - \alpha\beta\big)\big(x^2 + (\alpha - \beta)x - \alpha\beta\big)$$

has all coefficients in $\mathbf{K}$. It follows immediately that (ii) implies (i).

The basic relations among the coefficients and the roots are $\alpha^2 + \beta^2 = -b$ and $\alpha^2\beta^2 = d$. These together with $(\alpha \pm \beta)^2 = -b \pm 2\alpha\beta$ imply that $\beta^2$ or $\alpha\beta$ is in $\mathbf{K}$ if $\alpha^2$ or $\alpha \pm \beta$, respectively, is in $\mathbf{K}$. Hence (i) implies (ii).

The equivalence of (ii) and (iii) can be seen as follows. As consequences of the basic relations we obtain

(1) $(\alpha \pm \beta)^2 = -b \pm 2\sqrt{d}$

(2) $(b + 2\alpha^2)^2 = (\alpha^2 - \beta^2)^2 = (\alpha + \beta)^2(\alpha - \beta)^2 = b^2 - 4d$.

Hence $\alpha \pm \beta \in \mathbf{K}$ if and only if $-b \pm 2\sqrt{d} \in \mathbf{K}^2$. Similarly, $\alpha^2 \in \mathbf{K}$ is equivalent to $b^2 - 4d \in \mathbf{K}^2$. □

THEOREM 3. *Let $\mathbf{h}(x) = x^4 + bx^2 + d$ be irreducible over $\mathbf{K}$, a field of characteristic $\neq 2$. Let $\pm\alpha, \pm\beta$ be its roots and $\mathbf{F}$ its splitting field. Then*:

(i) $\mathrm{Gal}(\mathbf{F}/\mathbf{K}) \cong \mathbf{V} \leftrightarrow d \in \mathbf{K}^2 \leftrightarrow \alpha\beta \in \mathbf{K}$;

(ii) $\mathrm{Gal}(\mathbf{F}/\mathbf{K}) \cong \mathbf{C}_4 \leftrightarrow d(b^2 - 4d) \in \mathbf{K}^2 \leftrightarrow \mathbf{K}(\alpha\beta) = \mathbf{K}(\alpha^2)$;

(iii) $\mathrm{Gal}(\mathbf{F}/\mathbf{K}) \cong \mathbf{D}_4 \leftrightarrow d$ *and* $d(b^2 - 4d) \notin \mathbf{K}^2 \leftrightarrow \alpha\beta \notin \mathbf{K}(\alpha^2)$.

*Proof.* The cubic resolvent for $\mathbf{h}(x)$ is $\mathbf{r}(x) = (x - b)(x^2 - 4d)$. Thus Theorem 1 implies that the Galois group of $\mathbf{h}(x)$ is isomorphic to one of the three two-power groups. Again by Theorem 1, $\mathrm{Gal}(\mathbf{F}/\mathbf{K}) \cong \mathbf{V}$ if and only if $\mathbf{r}(x)$ splits over $\mathbf{K}$. This is clearly equivalent to $\sqrt{d} = \alpha\beta \in \mathbf{K}$, and (i) is established.

Now assume that $\sqrt{d} \notin \mathbf{K}$, so that $\mathbf{r}(x)$ has only one root, $t_1 = b$, in $\mathbf{K}$, and we have case (iv) or (v) of Theorem 1. The polynomial $\mathbf{g}(x)$ associated with $\mathbf{h}(x)$ is $\mathbf{g}(x) = (x^2 - bx + d)x^2$. The nonzero roots of $\mathbf{g}(x)$ are $-\alpha^2 = (1/2)(b + \sqrt{b^2 - 4d})$ and $-\beta^2 - (1/2)(b - \sqrt{b^2 - 4d})$. Since $\mathbf{h}(x)$ is irreducible these roots are not in $\mathbf{K}$. By Theorem 1, $\operatorname{Gal}(\mathbf{F}/\mathbf{K}) \cong \mathbf{C}_4$ if and only if $\mathbf{g}(x)$ splits in $\mathbf{E} = \mathbf{K}(\sqrt{d})$. This is the same as saying that $\mathbf{K}(\sqrt{b^2 - 4d}) = \mathbf{K}(\sqrt{d})$, or equivalently, $\mathbf{K}(\alpha^2) = \mathbf{K}(\alpha\beta)$. Two quadratic extensions of a field are equal if and only if the product of their discriminants is a square in the field, so the last condition is equivalent to $d(b^2 - 4d) \in \mathbf{K}^2$. This establishes (ii). Statement (iii) follows by mutual exclusion. □

**4. Examples.**

*Example* 1. The polynomials $\mathbf{p}_1(x) = x^4 - 4x^3 + 4x^2 + 6$ and $\mathbf{p}_2(x) = x^4 + 4x^3 + 6x^2 + 4x + 2$ are irreducible by Eisenstein's criterion. In the first case $\mathbf{r}_1(x) = x(x^2 - 4x - 24)$ and $\mathbf{E} = \mathbf{Q}(\sqrt{7})$. Now $\mathbf{g}_1(x) = (x - 2)^2(x^2 + 6)$ obviously does not split over $\mathbf{E}$, so $\operatorname{Gal}(\mathbf{F}/\mathbf{Q}) \cong \mathbf{D}_4$. In the second case $\mathbf{r}_2(x) = x(x - 2)(x - 4)$, hence $\operatorname{Gal}(\mathbf{F}/\mathbf{Q}) \cong \mathbf{V}$.

The polynomial $\mathbf{p}_3(x) = x^4 + x^3 + x^2 + x + 1$ is irreducible as the minimal polynomial for the primitive fifth roots of unity, and it follows from other considerations that $\operatorname{Gal}(\mathbf{F}/\mathbf{Q}) \cong \mathbf{C}_4$. This can be seen directly from Theorem 1 by observing that $\mathbf{r}_3(x) = (x - 2)(x^2 + x - 1)$ and $\mathbf{g}_3(x) = (x - 1)^2(x^2 + x - 1)$. Thus $\operatorname{Gal}(\mathbf{F}/\mathbf{Q}) \cong \mathbf{C}_4$, since $\mathbf{r}_3(x)$ and $\mathbf{g}_3(x)$ have the same splitting field.

*Example* 2. The polynomial $\mathbf{f}(x) = x^4 + px + p$ is irreducible by Eisenstein's criterion. Since its cubic resolvent $\mathbf{r}(x) = x^3 - 4px - p^2$ is an integer monic polynomial it follows that $\mathbf{r}(x)$ is irreducible if and only if it has no integer roots. Possible roots are the divisors of $-p^2$. We observe that $\mathbf{r}(1) = 1 - 4p - p^2$, $\mathbf{r}(-1) = -1 + 4p - p^2$, $\mathbf{r}(p) = p^2(p - 5)$, $\mathbf{r}(-p) = p^2(3 - p)$, $\mathbf{r}(p^2) = p^2(p^4 - 4p - 1)$, and $\mathbf{r}(-p^2) = p^2(-p^4 + 4p - 1)$. It follows that $\mathbf{r}(x)$ is irreducible if and only if $p \neq 3, 5$. This together with $D = -p^3(99p - 64) < 0$ implies that $\operatorname{Gal}(\mathbf{F}/\mathbf{Q}) \cong S_4$ if $p \neq 3, 5$.

If $p = 3$, then $\mathbf{r}(x) = (x + 3)(x^2 - 3x - 3)$ and $\mathbf{E} = \mathbf{Q}(\sqrt{21})$. Now $\mathbf{g}(x) = (x^2 + 3x + 3)(x^2 + 3)$ has no real roots, so it does not split over $\mathbf{E}$. Hence $\operatorname{Gal}(\mathbf{F}/\mathbf{Q}) \cong \mathbf{D}_4$.

If $p = 5$, then $\mathbf{r}(x) = (x - 5)(x^2 + 5x + 5)$ and $\mathbf{E} = \mathbf{Q}(\sqrt{5})$. Here we have $\mathbf{g}(x) = (x^2 - 5x + 5)(x^2 - 5)$ which splits over $\mathbf{E}$. Thus $\operatorname{Gal}(\mathbf{F}/\mathbf{Q}) \cong \mathbf{C}_4$.

*Example* 3. Let $k$, $j$ be different square-free integers and $0 \neq r, s \in \mathbf{Q}$. Then $\mathbf{u}(x) = x^4 - 2(r^2k + s^2j)x^2 + (r^2k - s^2j)^2$ is irreducible over $\mathbf{Q}$, since $b^2 - 4d = 16r^2s^2kj$, $-b + 2\sqrt{d} = 4r^2k$, and $-b - 2\sqrt{d} = 4s^2j$ are not squares in $\mathbf{Q}$. Theorem 3 implies that $\operatorname{Gal}(\mathbf{F}/\mathbf{Q}) \cong \mathbf{V}$, since $d = (r^2k - s^2j)^2 \in \mathbf{Q}^2$.

A normal extension of $\mathbf{Q}$ has Galois group isomorphic to $\mathbf{V}$ if and only if it can be written as $\mathbf{Q}(\sqrt{k}, \sqrt{j})$ for two distinct square-free integers $k$ and $j$. Since the roots $\pm(r\sqrt{k} + s\sqrt{j})$ and $\pm(r\sqrt{k} - s\sqrt{j})$ of $\mathbf{u}(x)$ lie in $\mathbf{Q}(\sqrt{k}, \sqrt{j})$, every normal extension with Galois group $\mathbf{V}$ can be obtained as a splitting field of a polynomial of this form.

*Example* 4. Let $m, n \in \mathbf{Z}$ with $m^2 + n^2 \notin \mathbf{Q}^2$. Then the polynomial $\mathbf{v}(x) = x^4 - 2(m^2 + n^2)x^2 + n^2(m^2 + n^2)$ is irreducible over $\mathbf{Q}$ since $m^2 + n^2 \notin \mathbf{Q}^2$ implies that $b^2 - 4d = 4m^2(m^2 + n^2)$ and $-b \pm 2\sqrt{d} = 2(m^2 + n^2) \pm 2n\sqrt{m^2 + n^2}$

are not squares in $\mathbf{Q}$. By Theorem 3 we obtain $\mathrm{Gal}(\mathbf{F}/\mathbf{Q}) \cong \mathbf{C}_4$, since $d(b^2 - 4d) = 4m^2n^2(m^2 + n^2)^2 \in \mathbf{Q}^2$.

*Example* 5. For any pair, $p, q$ of different odd primes, the polynomial $\mathbf{w}(x) = x^4 - px^2 + q$ is irreducible over $\mathbf{Q}$. This follows immediately by Theorem 2 if we can show $b^2 - 4d = p^2 - 4q \notin \mathbf{Q}^2$, since $p \pm 2\sqrt{q} \notin \mathbf{Q}^2$.

Without loss of generality we may assume $p^2 - 4q > 0$. Suppose $p^2 - 4q = t^2$ for some integer $t > 0$. Then $t$ is odd and each of the factors of $4q = (p + t)(p - t)$ is even. Because $q$ is prime and $0 < p - t < p + t$, we must have $p - t = 2$ and $p + t = 2q$. However, this implies that $p = q + 1$, contradicting the assumption that $p$ and $q$ are both odd.

Similarly, $d = q$ and $d(b^2 - 4d) = q(p^2 - 4q)$ are not squares in $\mathbf{Q}$; hence $\mathrm{Gal}(\mathbf{F}/\mathbf{Q}) \cong \mathbf{D}_4$ by Theorem 3. Since the most commonly used examples of polynomials with $\mathbf{D}_4$ as a Galois group have a pair of real and a pair of complex roots, we note that $\mathbf{w}(x)$ has all roots real if $p^2 - 4q > 0$.

**Acknowledgement.** The authors want to thank Frank Stephanic for his contribution to Theorem 1.

REFERENCES

1. G. Bucht, Über einige algebraischen Körper achten Grades, *Arkiv für Matematik, Astronomi o. Fysik*, 6, no. 30 (1910) 1–36.
2. R. Garver, Quartic Equations with Certain Groups, *Annals of Math.*, 30 (1928/29) 47–51.
3. F. Hack, Beiträge zur Anwendung der Gruppentheorie auf kubische und biquadratische Gleichungen, Dissertation, Tübingen, 1895.
4. O. Hölder, Galois'sche Theorie mit Anwendungen, Encyklopädie der Mathematischen Wissenschaften, I. 1, Arithmetik und Algebra, Teubner, Leipzig, 1898–1904, pp. 480–520.
5. T. W. Hungerford, Algebra, Holt, Rinehart and Winston, Inc., New York, 1974.
6. N. Jacobson, Basic Algebra I, W. H. Freeman and Company, San Francisco, 1974.
7. I. Kaplansky, Fields and Rings, second edition, University of Chicago Press, Chicago, 1972.
8. W. Krull, Elementare und klassische Algebra, Band II, Sammlung Göschen, Berlin, 1959.
9. F. Seidelmann, Die Gesamtheit der kubischen und biquadratischen Gleichungen mit Affekt bei beliebigem Rationalitäsbereich, *Math. Annalen*, 18 (1918) 230–233.
10. R. P. Stauduhar, Determination of Galois Groups, *Mathematics of Computation*, 27 (1973) 981–996.
11. B. L. van der Waerden, Modern Algebra, vol. 1, New York, Ungar Publishing Company, 1949.

# Comparing the Spectral Radii of Two Nonnegative Matrices

R. B. BAPAT
*Indian Statistical Institute, 7, S. J. S. Sansanwal Marg, New Delhi 110016, India*

If $A$ is an $n \times n$ matrix, then $r(A)$ will denote the spectral radius of $A$, which by definition is the maximum modulus of an eigenvalue of $A$. It is well known that if $A = ((a_{ij}))$ and $B = ((b_{ij}))$ are nonnegative $n \times n$ matrices such that $b_{ij} \geqslant a_{ij}$ for all $i, j$, then $r(B) \geqslant r(A)$. This is usually proved as a consequence of the Collatz-Wielandt characterization of the spectral radius of a nonnegative matrix (see, for example [**1**, pp. 27–28]). In this note we prove a more general result in an elementary way.

If $A$ is a nonnegative $n \times n$ matrix, then by the Perron–Frobenius theorem, $r(A)$ is an eigenvalue of $A$, corresponding to which there are nonnegative left and right eigenvectors. We refer to these as Perron eigenvectors. The result that we prove gives a bound for the ratio of the spectral radii of two nonnegative matrices in terms of the entries of the matrices and the components of the left and the right Perron eigenvectors of one of them. In what follows, we take $(z/0)^0 = 1$ for any real $z$.

THEOREM. *Let $A$, $B$ be nonnegative $n \times n$ matrices with $r(A) > 0$ and suppose $A$ admits left and right Perron eigenvectors $x$, $y$ respectively, with $\sum_{i=1}^n x_i y_i = 1$ (this is satisfied, for example, if $A$ is irreducible). Then*

$$\frac{r(B)}{r(A)} \geqslant \prod_{i,j=1}^{n} \left(\frac{b_{ij}}{a_{ij}}\right)^{a_{ij}x_i y_j / r(A)} \tag{1}$$

*Furthermore, if the $x$, $y$ are positive and if $B$ has a positive (right or left) Perron eigenvector, then equality holds in (1) if and only if there exist positive $\mu, \alpha_i$, $i = 1, 2, \ldots, n$ such that $b_{ij} = \mu a_{ij}\alpha_i/\alpha_j$, $i, j = 1, 2, \ldots, n$.*

*Proof.* We first prove (1) under the additional assumption that $B$ has a positive right Perron eigenvector which we denote by $v$ and we let $u_i = x_i y_i / v_i$, $i = 1, 2, \ldots, n$.

Let $K = \{(i, j): a_{ij} x_i y_j > 0\}$. We have,

$$\begin{aligned}
\prod_{i,j=1}^{n} \left\{\frac{b_{ij}u_i v_j / r(B)}{a_{ij}x_i y_j / r(A)}\right\}^{a_{ij}x_i y_j / r(A)} &= \prod_{(i,j)\in K} \left\{\frac{b_{ij}u_i v_j / r(B)}{a_{ij}x_i y_j / r(A)}\right\}^{a_{ij}x_i y_j / r(A)} \\
&\leqslant \sum_{(i,j)\in K} \frac{a_{ij}x_i y_j}{r(A)} \frac{b_{ij}u_i v_j / r(B)}{a_{ij}x_i y_j / r(A)} \\
&= \sum_{(i,j)\in K} \frac{b_{ij}u_i v_j}{r(B)} \\
&\leqslant \sum_{i,j=1}^{n} \frac{b_{ij}u_i v_j}{r(B)} \\
&= \sum_{i=1}^{n} u_i \sum_{j=1}^{n} \frac{b_{ij}v_j}{r(B)} \\
&= 1,
\end{aligned}$$

where the first inequality follows by the generalized (weighted) arithmetic mean–geometric mean inequality,

Hence,

$$\prod_{i,j=1}^{n} \left\{\frac{b_{ij}u_i v_j}{r(B)}\right\}^{a_{ij}x_i y_j / r(A)} \leqslant \prod_{i,j=1}^{n} \left\{\frac{a_{ij}x_i y_j}{r(A)}\right\}^{a_{ij}x_i y_j / r(A)} \tag{2}$$

Inequality (1) is obtained from (2) after a trivial simplification.

From the proof it is clear that equality holds in (1) if and only if for some positive $\lambda$,

$$\frac{b_{ij}u_i v_j}{r(B)} = \lambda \frac{a_{ij}x_i y_j}{r(A)}, \qquad i, j = 1, 2, \ldots, n \tag{3}$$

Now suppose $x, y$ are positive. Since $u_i = x_i y_i / v_i$ for all $i$, we have from (3),

$$b_{ij} = \lambda \frac{r(B)}{r(A)} a_{ij} \frac{v_i}{y_i} \frac{y_j}{v_j}, \qquad i, j = 1, 2, \ldots, n.$$

Set

$$\mu = \lambda \frac{r(B)}{r(A)}, \qquad \alpha_i = \frac{v_i}{y_i}, \qquad i = 1, 2, \ldots, n$$

and we get the assertion about equality in (1). The proof of (1) as well as that of the assertion about equality is similar when $B$ is assumed to have a positive left eigenvector.

To prove (1) for an arbitrary nonnegative matrix $B$, approximate $B$ by a sequence of positive matrices $B^{(k)}$, $k = 1, 2, \ldots$. Since positive matrices have positive Perron eigenvectors, each $B^{(k)}$ satisfies (1). Now observe that both sides of (1) are continuous in the entries of $B$ and the proof is complete.

REFERENCE

1. R. S. Varga, Matrix Iterative Analysis, Prentice Hall, New Jersey, 1962.

# THE TEACHING OF MATHEMATICS

EDITED BY MELVIN HENRIKSEN AND STAN WAGON

## From Calculus to Number Theory

JAMES DUEMMEL
*Department of Mathematics, Western Washington University, Bellingham, WA 98225*

While preparing an exam for a beginning calculus course, I decided to use the following problem. *If we have a sheet of metal a inches by b inches and cut identical squares of size x by x from the four corners then we can fold up the edges to form a box. Find that x which yields the maximum volume.* For simplicity I wanted to use integers $a$ and $b$ such that the $x$ yielding the maximum,

$$x = \frac{a + b - \sqrt{(a+b)^2 - 3ab}}{6},$$

would be rational. Since the problem is symmetric in $a$ and $b$ and very easy for a square, we will assume $a < b$.

Clearly $x$ will be rational if and only if $(a+b)^2 - 3ab$ is a perfect square. We want an integer $m$ such that

$$(a+b)^2 - 3ab = m^2$$

or

$$a^2 + b^2 - ab = m^2. \tag{1}$$

Readers familiar with the characterization of Pythagorean triples ([**1**, p. 242] or [**4**, p. 391]) will have already noted a similarity to the present situation. The problem is also discussed in [**3**], where the emphasis is on the boxes while here we will concentrate on the numbers $m$, $a$ and $b$.

Simple observations about common factors suggest that we can restrict ourselves to searching for solutions of (1) in which $a$ and $b$ are relatively prime.

To quickly find a solution for my exam I used a simple computer search which produced the following list.

| $m$ | $a$ | $b$ | $x$ | $m$ | $a$ | $b$ | $x$ |
|---|---|---|---|---|---|---|---|
| 7 | 3 | 8 | 2/3 | 43 | 13 | 48 | 3 |
| 7 | 5 | 8 | 1 | 43 | 35 | 48 | 20/3 |
| 13 | 7 | 15 | 3/2 | 49 | 16 | 55 | 11/3 |
| 13 | 8 | 15 | 5/3 | 49 | 39 | 55 | 15/2 |
| 19 | 5 | 21 | 7/6 | 61 | 9 | 65 | 13/6 |
| 19 | 16 | 21 | 3 | 61 | 56 | 65 | 10 |
| 31 | 11 | 35 | 5/2 | 67 | 32 | 77 | 7 |
| 31 | 24 | 35 | 14/3 | 67 | 45 | 77 | 55/6 |
| 37 | 7 | 40 | 5/3 | 73 | 17 | 80 | 4 |
| 37 | 33 | 40 | 6 | 73 | 63 | 80 | 35/3 |

While this list together with multiples of its entries probably provides a lifetime supply of "nice" calculus problems, it also suggests some interesting questions.

If $(m, a, b)$ satisfies (1) with $a < b$ does the triple $(m, b - a, b)$ also satisfy (1)? Is $m \equiv 1 \pmod 6$ always true? 25 and 55 satisfy $m \equiv 1 \pmod 6$ but neither appears in the list. Is no $m$ a multiple of 5?

The answer to all three questions is "yes." The first involves simple algebra and has an interesting geometric interpretation involving the law of cosines and an equilateral triangle divided by a line through one of its vertices into two triangles with integral sides (see [**2**]).

DEFINITION. *A triple of positive integers $(m, a, b)$ is said to be a box triple if $m^2 = a^2 + b^2 - ab$ and $a < b$. A box triple $(m, a, b)$ is primitive if $2a < b$ and $a$ and $b$ are relatively prime.*

In a box triple any pair of the three numbers $m, a, b$ is relatively prime if and only if the only common divisor of all three is 1. We will deal with primitive box triples.

The second and third questions raised above require some work. First we note $m$ must be odd because $a$, $b$ and $m$ are relatively prime. If we multiply by 4 and complete the square in (1) we obtain $(2b - a)^2 + 3a^2 = 4m^2$ and hence

$$3a^2 = (2m + 2b - a)(2m - 2b + a).$$

Note that the sum and difference of the factors on the right are $4m$ and $2(2b - a)$, respectively. The numbers $4m$ and $2(2b - a)$ cannot have an odd factor other than 1 in common with $a$ since then $a$, $b$ and $m$ would have a common factor other than 1.

It follows that $2m - 2b + a$ and $2m + 2b - a$ have no *odd* prime factors in common and that, with the possible exception of 3, any odd prime divisor of $2m - 2b + a$ or $2m + 2b - a$ is a divisor of $a$. By grouping the odd prime divisors of $a$ according to whether they divide $2m - 2b + a$ or $2m + 2b - a$, we find *odd* numbers $r$ and $s$ and integers $K$, $T$ and $n$ such that

$$\begin{aligned} \gcd(3r, s) = 1 \qquad K + T &= 2n \\ a &= 2^n rs \\ 2m - 2b + a &= 3 \cdot 2^K r^2 \\ 2m + 2b - a &= 2^T s^2 \end{aligned} \tag{2}$$

(or a similar case with the right sides of the last two equations reversed).

We need to know more about $K$ and $T$. Note that their sum must be even. From the last two equations in (2) we see

$$4m = 2^T s^2 + 3 \cdot 2^K \cdot r^2, \tag{3}$$

and we previously established that $m$ is odd. Even/odd arguments for various cases eliminate all possibilities for $K$ and $T$ except (i) $K = T = 0$, and (ii) one of $K$ and $T$ is 2 and the other is an even integer greater than 2.

Let $K = 2k$, $T = 2t$. Then (3) may be rewritten as

$$4m = 4^t s^2 + 3 \cdot 4^k \cdot r^2. \tag{4}$$

In this equation, if the arithmetic is performed modulo 3, we have

$$m \equiv s^2 + 3r^2 \equiv s^2 \pmod 3.$$

The only possible values of $s^2$ modulo 3 are 0 and 1. If $s^2 \equiv 0 \pmod 3$, then $3|s$. This contradicts $\gcd(3r, s) = 1$. Hence $m \equiv s^2 \equiv 1 \pmod 3$.

Since $m$ is an odd number, this implies $m \equiv 1 \pmod 6$. We have an affirmative answer to the second question raised above.

Using (4) and arithmetic modulo 5 we can show that $m$ cannot be divisible by 5, thereby answering the third question. But we will obtain a stronger result later. With affirmative answers to the second and third questions we have found necessary conditions that a number be an $m$ in a primitive triple. The converse question naturally arises. Are the conditions $m \equiv 1 \pmod 6$ and $m \not\equiv 0 \pmod 5$ sufficient to make a number an $m$ in a primitive box triple?

This question prompted another computer search. Here are the next twenty entries in the list of primitive box triples.

| $m$ | $a$ | $b$ | $x$ | $m$ | $a$ | $b$ | $x$ |
|---|---|---|---|---|---|---|---|
| 79 | 40 | 91 | 26/3 | 151 | 56 | 171 | 38/3 |
| 91 | 19 | 99 | 9/2 | 157 | 25 | 168 | 6 |
| 91 | 11 | 96 | 8/3 | 163 | 75 | 187 | 33/2 |
| 97 | 55 | 112 | 35/3 | 169 | 15 | 176 | 11/3 |
| 103 | 40 | 117 | 9 | 181 | 104 | 109 | 22 |
| 109 | 24 | 119 | 17/3 | 193 | 32 | 207 | 23/3 |
| 127 | 13 | 133 | 19/6 | 199 | 56 | 221 | 13 |
| 133 | 65 | 153 | 85/6 | 211 | 29 | 224 | 7 |
| 133 | 23 | 143 | 11/2 | 217 | 87 | 247 | 39/2 |
| 139 | 69 | 160 | 15 | 217 | 17 | 225 | 25/6 |

There are some surprises in the list. The numbers 121 and 187 do not appear as $m$s although they satisfy both conditions $m \equiv 1 \pmod 6$ and $m \not\equiv 0 \pmod 5$. Other numbers such as 91, 133 and 217 appear as $m$s in *two* distinct primitive triples.

A longer search revealed many similar occurrences. Clearly the conditions $m \equiv 1 \pmod 6$ and $m \not\equiv 0 \pmod 5$ are not sufficient to assure that $m$ will appear as an $m$ in a primitive box triple. And why do some $m$s, but not others, appear in more than one primitive triple?

The key to these questions seems to be in the prime factorization of the $m$s. For example, 121, 187, and the next missing $m$, 257, have prime factorizations $121 = 11^2$, $187 = 11 \cdot 17$, and $257 = 23 \cdot 11$. Each contains prime factors *not* congruent to 1 modulo 6. For those numbers which appear in several primitive triples we find $91 = 7 \cdot 13$, $133 = 7 \cdot 19$, $217 = 7 \cdot 31$ in which all of the primes are congruent to 1 modulo 6. All the remaining $m$s in our lists are primes or squares of primes congruent to 1 modulo 6.

The numerical evidence strongly suggests a new conjecture: An integer is an $m$ in a primitive box triple if and only if all of its prime factors are congruent to 1 modulo 6.

We will prove the necessity of the condition. Suppose that $m$ comes from a primitive box triple and in (4) set $S = 2^t s$ and $R = 2^k r$. Rewrite (4) as

$$3R^2 + S^2 = 4m. \tag{5}$$

Let $p$ be any odd prime divisor of $m$. $p$ cannot divide both $R$ and $S$ since $\gcd(3r, s) = 1$. Then by (5) it cannot divide either of them. There exists some integer $c$ such that $Rc \equiv 1 \pmod p$. Then (5) implies $3 + (Sc)^2 \equiv 0 \pmod p$ or, with $x = Sc$,

$$x^2 \equiv -3 \pmod p;$$

$-3$ must be a square, a quadratic residue, modulo $p$.

To this point our mathematical tools have been only elementary properties of prime numbers, congruences and the "divides" relation. But now we turn to a deeper classical result in number theory, the law of quadratic reciprocity ([**1**, Chap. 9] or [**4**, Chap. 9]). From this famous result the following lemma follows easily.

LEMMA 1. *For an odd prime p larger than* 3 *the equation* $x^2 \equiv -3$ *modulo p has a solution if and only if* $p \equiv 1 \pmod 6$.

In view of the discussion that precedes Lemma 1 we can conclude that, for an $m$ in a primitive box triple, every prime factor of $m$ must be congruent to 1 modulo 6 and the necessity part of the following theorem has been verified.

THEOREM. *An integer is an m in a primitive box triple if and only if all of its prime factors are congruent to* 1 *modulo* 6.

The proof of the sufficiency that we found is lengthy and detailed. So we invite the reader to construct a proof. Crucial to our proof was the following pair of identities.

$$(3x^2 + y^2)(3u^2 + v^2) = 3(xv + yu)^2 + (3xu - yv)^2.$$

$$(3x^2 + y^2)(3u^2 + v^2) = 3(xv - yu)^2 + (3xu + yv)^2.$$

We should note that another topic in number theory arises here, the famous theorem of Lagrange ([**1**, p. 279]) which states that every positive integer is the sum of at most four sqaures. In our problem we dealt with the special case $4m = s^2 + 3r^2$ in which $4m$ is written as the sum of four squares of which three are the same. One result along the way in the sufficiency proof is this: Every prime congruent to 1 mod 6 is of the form $3x^2 + y^2$ for some positive integers $x$ and $y$.

REFERENCES


1. David M. Burton, Elementary Number Theory, Allyn and Bacon, Boston, MA, 1980.
2. Kay Dundas, Quasi-Pythagorean triples for an oblique triangle, *The Two-year College Mathematics Journal*, 8 (1977) 152–155.
3. Kay Dundas, To build a better box, *The College Mathematics Journal*, 15 (1984) 30–36.
4. Kenneth H. Rosen, Elementary Number Theory and its Applications, Addison Wesley, Reading, MA, 1984.


## A Note on The Row-Reduction Algorithm


CHIH-HAN SAH

*Department of Mathematics, SUNY, Stony Brook, NY 11794*


Classically, the Gaussian Elimination Method is used to solve a system of $n$ linear equations in $n$ unknowns (see [**2**, §302] for related methods). By setting some of the coefficients to 0, there is no problem extending the method to the general case

of $m$ equations in $n$ unknowns:

$$\begin{aligned} a_{1,1}x_1 + \cdots + a_{1,n}x_n &= b_1 \\ &\vdots \\ a_{m,1}x_1 + \cdots + a_{m,n}x_n &= b_m. \end{aligned} \tag{1}$$

Quite frequently, the above system is abbreviated to the matrix equation:

$$AX = B, \quad A \text{ is } m \times n, \quad X \text{ is } n \times 1, \quad \text{and } B \text{ is } m \times 1. \tag{2}$$

With the usual convention on row and column indices, the method is often called the Row-Reduction Algorithm in elementary texts. Depending on the level of the text, the matrix equation (2) may be explained in terms of a linear transformation $A: F^n \to F^m$, where $F^r$ denotes the $F$-vector space of all column vectors with $r$ entries from the field $F$ of scalars. When $B$ is the zero column vector, (2) may also be explained in terms of testing for the $F$-linear independence of the $n$ columns of $A$ in $F^m$. Such explanations are usually accompanied by discourses on the concepts of im $A$, ker $A$, rank($A$), and nullity($A$).

When one finally gets around to solving (1), practically all the elementary texts apply the Row-Reduction Algorithm to the augmented matrix $(A, B)$. It is then necessary to explain the decoding process that yields the general form of the solution of (1) from the Row-Echelon Normal Form of $(A, B)$. More often than not, the students have lost their way by this time. Out of desperation, teachers frequently try to assure the students that the Row-Reduction Algorithm may be implemented on a desk-top personal computer even though other techniques may be used (the choice may depend on the field $F$ of scalars). Aside from the irritation of having to perform the decoding step by hand, the question of finding a basis for im $A$ is ignored. Also overlooked in this process is the fact that the original augmented matrix $(A, B)$ cannot be recovered from its row-echelon normal form without peeking into the memory bank of the computer.

As long as the modern trimmings of vector space and computer implementation are mentioned, the shortcomings mentioned above can be overcome with very little additional work. We can in fact let the computer do all the work for us. This will be described below.

To justify using a computer, one might as well explain that the system (1) may have to be solved over and over again where the single column vector $B$ is replaced by a finite set of column vectors in (2). To incorporate this possibility, we simply replace the two appearances of 1 in (2) by the positive integer $s$. In general, we use $R_{*j}$ to denote the $j$th column of the matrix $R$; the transpose of the matrix $R$ will be denoted by ${}^tR$; $I_u$ and $O_{v,w}$ denote the identity and the zero matrices of appropriate sizes, respectively. Instead of applying the row-reduction algorithm to $(A, B)$, we apply it to the following $(s + n) \times (s + m + n)$ matrix (in block form):

$$\begin{pmatrix} I_s & -{}^tB & O_{s,n} \\ O_{n,s} & {}^tA & I_n \end{pmatrix}. \tag{3}$$

Let us now assume that the row-echelon normal form of (3) is:

$$\begin{pmatrix} I_s & {}^tY & {}^tP \\ O_{n,s} & {}^tN & {}^tG \end{pmatrix}. \tag{4}$$

We then have the following "recipes":

$$\begin{array}{ll} \text{(a)} & AX_{*j} = B_{*j} \text{ has a solution if and only if } Y_{*j} = 0, \text{ where } 1 \leqslant j \leqslant s. \\ \text{(b)} & \text{The (right) column rank of } A \text{ is the number } r \text{ of nonzero rows of } {}^tN \text{ and } N_{*1}, \dots, N_{*r} \text{ is a basis for im } A. \\ \text{(c)} & \text{The (right) nullity of } A \text{ is } n - r \text{ and } G_{*r+1}, \dots, G_{*n} \text{ is a basis for } \ker A. \\ \text{(d)} & \text{If } Y_{*j} = 0, \text{ then the most general solution of } AX_{*j} = B_{*j} \text{ is } X_{*j} = P_{*j} + G_{*r+1}c_{r+1} + \cdots + G_{*n}c_n, \text{ where } c_k \text{ range over } F \text{ independently.} \end{array} \tag{5}$$

In order to explain to the more inquisitive students why the recipe works, one has to remind the students of the following items: row operations correspond to left multiplication by invertible matrices; row-reduction involves specific orderings of the rows and the columns in addition to a specified convention on leading terms of the nonzero rows (in particular, the vanishing of the entries in the column above the leading entries).

After the preceding reminders, it is then clear that the introduction of the "split-up" identity matrix $I_{s+n}$ in (3) forces us to reach (4) from (3) through left multiplication of (3) by the following invertible matrix:

$$\begin{pmatrix} I_s & {}^tP \\ O_{n,s} & {}^tG \end{pmatrix}.$$

Now, (a) of (5) follows because the translation of $Y_{*j} = 0$ through the row-reduction algorithm (applied to (3)) is precisely the statement that $B_{*j}$ is a suitable linear combination of the columns of $A$.

Matrix multiplication shows that ${}^tG{}^tA = {}^tN$. Since $G$ is invertible, it is immediate that the columns of $A$ and $N$ span the same subspace of $F^m$. It is also clear that ${}^tN$ is the row-echelon normal form of ${}^tA$ because of the presence of $I_s$ and $O_{n,s}$ in (3). These yield (b) of (5).

By taking tranposes, we have $AG = N$. By the choice of $r$, the last $n - r$ columns of $N$ are zeroes so that the last $n - r$ columns of $G$ lie in $\ker A$. Since $G$ is invertible, these last $n - r$ columns of $G$ must be linearly independent. We thus obtain (c) of (5).

One more matrix multiplication shows that ${}^tY = -{}^tB + {}^tP{}^tA$. By taking the transpose and looking at the $j$th column, (d) then follows from (c) in (5).

If we wish, we could now say that the "absent-minded boss" wants to know the original equation (2) from (4). To answer this question, we simply exchange the second and the third "column blocks" in (4) and apply the row-reduction algorithm to the result. The outcome is simply (3) with its second and third "column blocks" exchanged.

As an example, consider the equation $x + 2y = 3$. The Row-Reduction Algorithm takes on the following form:

$$\left(\begin{array}{c|c|cc} 1 & -3 & 0 & 0 \\ \hline 0 & 1 & 1 & 0 \\ 0 & 2 & 0 & 1 \end{array}\right) \to \left(\begin{array}{c|c|cc} 1 & 0 & 0 & 3/2 \\ \hline 0 & 1 & 0 & 1/2 \\ 0 & 0 & 1 & -1/2 \end{array}\right).$$

The most general solution is then given by

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 3/2 \end{pmatrix} + \begin{pmatrix} 1 \\ -1/2 \end{pmatrix} c.$$

The $1 \times 1$ column vector (1) evidently does span the column space of $A = (1,2)$. The other claims may be checked easily.

It is apparent from our preceding discussions that the use of the transpose could have been avoided if we had employed a "Column-Reduction Algorithm" in place of the customary Row-Reduction Algorithm. (Of course, one then writes up and down from right to left; in particular, one multiplies from the right. This sits well with a class that has some students brought up in the appropriate oriental tradition.)

The use of the transpose can be exploited as a setting to explain the concept of the dual space. In particular, one can extend the preceding discussion to include solving (1) over a division ring $D$. In this generality, it is essential that $D^m$ and $D^n$ in (2) be viewed as *right* $D$-vector spaces. The $D$-dual of these are naturally *left* $D$-vector spaces which can then be identified with spaces of row vectors. In this respect, row reduction of $A$ (from the left) and column reduction of $A$ (from the right) are quite different. As a good exercise, students may be asked to show that the *left* row rank of $A$ coincides with the *right* column rank of $A$. To emphasize this point, the following example over the real quaternion division algebra can be cited to show that the right column rank of $A$ need not be the same as its right row rank:

$$\begin{pmatrix} 1 & j \\ i & k \end{pmatrix}.$$

In terms of using the transpose ${}^tR$ of a matrix $R$ over a division ring $D$ to solve (1) as described, it is essential to point out that the row-reduction of ${}^tR$ must correspond to column reduction of $R$. This means that the entries of ${}^tR$ have to be viewed as lying in the opposite ring $D^{op}$ of $D$ (sometimes called the mirror image of $D$) so that left multiplication of the rows of ${}^tR$ really corresponds to right multiplication if ${}^tR$ is viewed to have entries from $D$. The preceding example may be used to exemplify the needed modification. With the dual spaces at hand, it is then apparent that the classical Gaussian Elimination Method is a manipulation in the dual space while the determination of the column vector $B$ in (2) as a linear combination of the columns of the matrix $A$ is a question in the column vector space $F^m$ rather than its dual. The column-reduction algorithm (disguised in terms of the row-reduction algorithm as described) tackles this latter question directly while the Gaussian Elimination Method and the decoding process involve a sort of detour via "double duality."

We note finally that **[1]** contains an elegant treatment on solving (1) without getting involved with reduction algorithms in their explicit form.

REFERENCES

1. E. Artin, Galois Theory, Notre Dame, 1971.
2. Encyclopedic Dictionary of Mathematics, 2nd edition, MIT Press, Cambridge, 1987.

# On Markov Processes in Elementary Mathematics Courses

JOHN T. BALDWIN[1]

*Department of Mathematics, Statistics and Computer Science, University of Illinois, Chicago, IL 60680*

I give here a survey of the major results about Markov processes which are treated in many finite mathematics courses. These notes are the result of conversations with a number of colleagues and a certain amount of literature search. They delineate carefully the distinctions between the existence and uniqueness of fixed points and asymptotic fixed points of Markov processes. My primary motivation for writing up this material was the confusion about these distinctions that I found in a number of elementary texts. As noted below these results can be given 'high-powered' explanations. Such explanations can provide a valuable perspective; they can also unnecessarily intimidate. The arguments in this paper are all easily accessible to undergraduates.

A *Markov process* is a system consisting of a finite number of *states* $s_1, \dots, s_n$ and for each $i, j \leqslant n$ a probability $p_{i,j}$ of the system passing from state $i$ to state $j$. A *state vector* $\langle p_1, p_2, \dots, p_n \rangle$ is a list whose $i$th entry is the probability $p_i$ that the system is in state $i$. If there are $n$ states there is an $n \times n$ matrix $P$ of transition probabilities such that $\bar{x}P$ is the state vector after the process has been applied to the system in state $\bar{x}$. Each state vector and each row of $P$ is a *probability* or *stochastic vector*; that is, each entry is nonnegative and the sum of the entries is 1. We use $P$ to refer to either a Markov process or its associated matrix.

The following example illustrates these notions. After the first 11 games of the 1987 baseball season an examination of the results of these games showed that the St. Louis Cardinals had a 2/3 chance of winning the next game after a win but only a 1/2 chance of winning the next game after a loss. We can describe this situation by a Markov process with two states $W$ (win) and $L$ (loss) and the following transition matrix:

$$\begin{pmatrix} 2/3 & 1/3 \\ 1/2 & 1/2 \end{pmatrix}.$$

After these 11 games the team had a winning percentage of .636 and a losing percentage of .364.

Does the information about the probability of winning after a loss provide a different (better) predictor of the percentage of games the team will win in 1987 than assuming the 11 games were typical and predicting .636? The discussion that follows provides such a prediction; whether the model is applicable presents a more subtle problem.

An *equilibrium* or *steady-state vector* for the process $P$ is a stochastic vector $\bar{x}$ such that $\bar{x}P = \bar{x}$. Clearly, if the process reaches a steady state it remains there. So a steady state vector provides a solution to the problem of the previous paragraph. But is there such a vector and does the system ever enter the steady state? Can the

[1]This article was written during a visit to the University of Notre Dame in 1986/87.

answer to the last question depend on the stochastic vector to which one begins to apply the process?

If repeated iteration of the Markov process applied to any state vector yields a sequence of vectors that approach the unique equilibrium vector of $P$, we say $P$ is *asymptotically stable*. In this case by applying the iterates of $P$ to the basis vectors we see that the powers $P^n$ of an asymptotically stable Markov process tend to a limit matrix $P^\infty$.

We now consider three distinct but easily confused questions. When does a process $P$ have an equilibrium vector? When does a process $P$ have a unique equilibrium vector? When is a process $P$ asymptotically stable? The answers to these questions constitute various forms of the fundamental theorem of Markov processes.

I. WEAK FUNDAMENTAL THEOREM. *Every Markov process P has an equilibrium vector.*

A Markov Process is *regular* if in some power of the transition matrix $P$ every entry is strictly positive. We say a matrix $Q$ is *stationary* if all rows of $Q$ are the same. It is easy to see that each row of a stationary matrix $Q$ is an equilibrium vector for $Q$ and that a matrix $P$ is asymptotically stable if and only if $P^\infty$ exists and is stationary.

II. FUNDAMENTAL THEOREM. *Suppose that P is the transition matrix of a regular Markov chain.*

(i) *There is a unique stochastic vector* $\bar{x}$ *satisfying* $\bar{x}P = \bar{x}$.

(ii) *Moreover, for any stochastic vector* $\bar{a}$ *the sequence* $\bar{a}, \bar{a}P, \bar{a}P^2, \ldots$ *tends to* $\bar{x}$. *That is, P is asymptotically stable.*

The following proof was devised by Victor Harnik and me. It shows that the purely algebraic requirements of Theorem I and the first part of Theorem II can be obtained by 'elementary methods.' That is, the proof proceeds via the techniques taught in a finite mathematics course (although we induce on the size of the matrix).

*Proof of Theorems I and II.* i). We must show first that the equation $\bar{x}P = x$ has a solution that is a stochastic vector. It suffices to show that the equation $\bar{x}(P - I) = 0$ has a nontrivial solution in which all entries are nonnegative since we can divide by the length of that vector to obtain a stochastic solution.

There are two cases. Suppose first that one (or more) rows of $P$ has a 1 on the diagonal. If $p_{i,i} = 1$ then the $i$th coordinate vector is an equilibrium vector for $P$. For the second case, suppose that no diagonal entry of $P$ is 1. This implies $P - I$ satisfies the following conditions.

a) Each entry on the diagonal is negative.

b) All other entries are nonnegative. (If all entries in $P$ are positive all nondiagonal entries $p_{i,j}$ of $P - I$ satisfy $0 < p_{i,j} < 1$.)

c) The sum of each row is 0.

Now we perform column operations on $P$ to put $P$ in lower diagonal form with all entries off the diagonal nonnegative and all entries on the diagonal nonpositive. If we achieve this form with $r$ columns of 0's then each solution of the system of equations is a positive linear combination of $r$ parameters which we can take to be 1. Dividing the resulting vector by its length we obtain the required stochastic solution.

To obtain this form we work row by row changing all entries to the right of the diagonal to 0. For the first row, add $p_{1,i}/-p_{1,1}$ times the first column to the $i$th column. We will now show that after performing these operations the matrix obtained by deleting the first row and the first column satisfies conditions a) through c). Thus, we can complete the argument by induction. For this, let the matrix resulting from these operations be $Q$. That is,

$$q_{i,j} = p_{i,j} - \frac{p_{1,i}}{p_{1,1}} p_{i,1}.$$

Since the row sums of $P$ were 0, $-p_{i,i} \geqslant p_{i,1}$ and $-p_{1,1} \geqslant p_{1,i}$. Thus $(-p_{1,1})(-p_{i,i}) \geqslant p_{i,1}p_{1,i}$. Whence

$$(-p_{i,i}) \geqslant \frac{p_{i,1}p_{1,i}}{-p_{1,1}} \quad \text{and} \quad q_{i,i} \leqslant 0$$

yielding condition a). Condition b) is immediate since except in the first row we always added a positive number. For condition c) we compute the sum of the $i$th row of $Q$. To derive the third line below from the second recall that since the sum of the first row of $P$ is 0, $\Sigma_{j=2}^{n} p_{1,j} = -p_{1,1}$.

$$\begin{aligned}
\sum_{j=1}^{n} q_{i,j} &= p_{i,1} + \sum_{j=2}^{n} \left( p_{i,j} - \frac{p_{1,j}}{p_{1,1}} p_{i,1} \right) \\
&= \sum_{j=1}^{n} p_{i,j} - \frac{p_{i,1}}{p_{1,1}} \sum_{j=2}^{n} p_{1,j} \\
&= \sum_{j=1}^{n} p_{i,j} - \frac{p_{i,1}}{p_{1,1}}(-p_{1,1}) \\
&= 0 + p_{i,1} \\
&= p_{i,1}.
\end{aligned}$$

Since $q_{i,1} = p_{i,1}$, $\Sigma_{j=2}^{n} q_{i,j} = 0$ as required. This completes the proof of Theorem I.

For Theorem II.i) consider a regular matrix $P$. Then for some $m$ all entries in $P^m$ are positive. If $P^m$ has a unique fixed point, then $P$ has at most one fixed point. By the first part of the argument $P$ has at least one fixed point. So without loss of generality we can assume all entries in $P$ are positive (i.e., we work with $P^m$). Now in performing the column operations described above no diagonal entry can become 0 except the lower right-hand corner. For, the sum of the remaining entries in that row would then be 0, but they are all positive. Thus we first reach a column of 0's from the situation where the last two columns contain $n-2$ 0's followed by $-a$, $a$ and $b$, $-b$, respectively, where $a$ and $b$ are positive. Thus the solution space to $\bar{x}(P - I) = 0$ has dimension 1 and the system has a unique stochastic solution. □

The essential fact about a regular matrix $P$ is that its unique stable vector satisfies ii) of Theorem II. That is, $P$ is asymptotically stable in the sense defined above. I reproduce here a proof of this theorem that uses concepts from advanced calculus. The following argument was suggested by William Howard to simplify a version using the sup norm. Several of my conversations elicited the comment, "Theorem II.ii) holds because a regular map is contractive." In fact, this remark is correct if you deal with the correct metric on the correct space. A regular matrix need not, however, be contractive viewed as an operator on $\mathbf{R}^n$.

*Proof of Fundamental Theorem II.* ii). Let $v_0$ be the unique stochastic fixed point for $P$ (whose existence is guaranteed by Theorem II.i). Let $H$ be the hyperplane of stochastic vectors. Consider the subset $W = \{\bar{v} - \bar{v}_0 \colon \bar{v} \in H\}$. Let the *norm* of a row vector $\bar{y}$ with $n$ entries be

$$\|\bar{y}\| = \sum_{i \leqslant n} |y_i|.$$

We will show that if all entries of a stochastic matrix $Q$ are positive and $c$ is the smallest entry in $Q$ then for any $\bar{w} \in W$,

$$\|\bar{w}Q\| \leqslant |1 - c|\,\|\bar{w}\|.$$

Let $\bar{b}_j = \langle b_1, \dots b_n \rangle$ be the (transpose of the) $j$th column of $Q^n$ and for any $d$ let $\bar{d}$ denote the vector that is $d$ in each coordinate. Now since the sum of the components of $\bar{w}$ is 0

$$\bar{w}Q = \sum_j \bar{w} \cdot \bar{b}_j = \sum_j \bar{w} \cdot \left(\bar{b}_j - \bar{c}\right).$$

So

$$\|\bar{w}Q\| \leqslant \sum_j \|\bar{w}\| \left\|\left(\bar{b}_j - \bar{c}\right)\right\| \leqslant \sum_j \|\bar{w}\|\,\|\bar{1} - \bar{c}\| \leqslant |1 - c|\,\|\bar{w}\|. \tag{1}$$

Now an easy induction shows that for any $\bar{w} \in W$, $\|\bar{w}Q^n\| \leqslant |1 - c|^n \|\bar{w}\|$. Since $c$ is clearly less than $1/2$ this shows that the sequence $\|\bar{w}Q^n\|$ converges geometrically to 0. But if $\|(\bar{v} - \bar{v}_0)Q^n\|$ tends to 0 and $\bar{v}_0$ is fixed by $Q$, the sequence $\bar{v}Q^n$ tends to $\bar{v}_0$ as required. This completes the proof if all terms of $Q$ are positive. Thus, if $P$ is regular we have established the result for $Q = P^m$ (for some $m$). Now a calculation like (1), but easier, shows that for any stochastic $P$ and any $\bar{w} \in W$,

$$\|\bar{w}P\| \leqslant \|\bar{w}\|.$$

Thus the sequence $\|\bar{w}P^n\|$ converges to 0 since it is nonincreasing and contains a subsequence that converges to 0. □

Thus regularity is a sufficient condition for a Markov process $P$ to be asymptotically stable. But there are asymptotically stable processes that are not regular. Consider for example the matrix

$$\begin{pmatrix} 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix}.$$

On the basis of Theorems I and II.i) one might confuse the existence of a unique stochastic fixed point of $P$ with the condition that $P$ is regular. Consider the two state Markov state such that the probability of going from state 1 to state 2 is 1 and vice versa. Then $P$ has a unique stochastic fixed point $(1/2, 1/2)$. But it is easily seen that the matrix $P$ is not regular and is not asymptotically stable.

To characterize those Markov processes that are asymptotically stable requires some more terminology. Two complementary notions are involved: irreducibility and aperiodicity.

Define the relation '$j$ is accessible from $i$,' written $i \to j$, to mean that it is possible to move from state $i$ to state $j$ by repeatedly performing the Markov process. Thus $i \to j$ if and only if the $(i, j)$ entry is not zero in some power of $P$. Now partition the states of the process according to mutual accessibility. We will

say that the process $P$ is *irreducible* if it has only one accessibility class. There is an important distinction that allows us to characterize the irreducible Markov processes that are asymptotically stable. The process is irreducible if for each pair $i$ and $j$ there is a power $n$ such that $p_{i,j}^n > 0$. It is regular if there is a single $n$ such that $p_{i,j}^n > 0$ for all $i$ and $j$. The *period* of a state $i$ is the greatest common divisor of the $k$ such that $p_{i,i}^k > 0$. If the period is one we say the state is *aperiodic*. The next lemma (Lemma 1.2 and Theorem 1.4 of [7]) first allows us to extend these definitions from states to irreducible processes and then explains the role of regular processes in the irreducible case. On the basis of part a) of the lemma we call an irreducible Markov process *periodic* if each state is periodic and *aperiodic* otherwise. This result requires a straightforward manipulation of greatest common divisors. The lemma is proved in [5] arguing only with states and in [7] using the properties of matrix multiplication. The proof is probably inappropriate for the usual course in finite mathematics but would fit well in a course in applied matrix theory.

LEMMA. *Let $P$ be an irreducible Markov process.*
a) *All states of $P$ have the same period.*
b) *$P$ is aperiodic if and only if $P$ is regular.*

Now consider a possibly reducible Markov process. We will say an accessibility class of states is an *absorbing class* if no state outside the class is accessible to any member of the class. Note that by decomposing the set of states into maximal sets of mutually accessible states it is possible to determine the absorbing classes of the process and to determine which of them are irreducible. If the Markov process has more than one absorbing class the process is not asymptotically stable. Moreover, if the restriction to the absorbing class is periodic, the process is not asymptotically stable. In the only remaining case, the restriction to the absorbing class is regular. In this case $P$ is asymptotically stable. The following theorem summarizes this situation. It is proved by combining the analysis of relations between states with a computation of the limit of $P^n$ as in the proof of Theorem IV. Thus, we refer to Theorem 4.7 of [7] for the proof.

III. THEOREM. *Let $P$ be a Markov process with a unique absorbing class. A suitable ordering of the rows of $P$ yields the following canonical form*

$$\begin{pmatrix} R & 0 \\ S & Q \end{pmatrix}. \tag{2}$$

*If the restriction to the absorbing class is regular then $R$ is regular and the Markov process $P$ is asymptotically stable. Otherwise $P$ is not asymptotically stable.*

The distinction between a periodic and an asymptotically stable Markov process can be expressed solely in terms of linear algebra. A Markov process is asymptotically stable if the eigenspace of the eigenvalue 1 has dimension 1 and there are no other eigenvalues with absolute value 1. It is difficult for students who are unfamiliar with complex numbers to check this condition since there are Markov chains with complex eigenvalues of modulus one. It is easy to construct such an example with four states by means of the state diagram.

Note that if the Markov process $P$ has a unique equilibrium vector then it has a unique absorbing class.

Larry Lambe suggested a relatively easy proof of a further typical topic from finite mathematics courses. An *absorbing state* of a Markov chain is one that once entered cannot be escaped. It corresponds to a 1 on the diagonal of the associated matrix. A state that is not absorbing is said to be *transient*. An *absorbing Markov process* is one with at least one absorbing state such that each absorbing state is accessible from each transient state. (That is, if state $i$ is transient and state $j$ is absorbing, in some power of $P$ the $(i, j)$ entry is positive.)

IV. ABSORBING PROCESS THEOREM. *The matrix $P$ of an absorbing Markov process can be put in the following form*

$$\begin{pmatrix} I & 0 \\ R & Q \end{pmatrix},$$

*where $I$ is an identity matrix and $I - Q$ has an inverse $N$. The rows of $N \cdot R$ give the probability of eventually passing from the transient state indexed by the row of $N$ to the absorbing state indexed by the column of $R$.*

*Proof.* The desired form for $P$ is obtained by listing the absorbing states first. The idea of the proof is to compute $P^\infty$. It is easy to see that

$$P^n = \begin{pmatrix} I & 0 \\ (I + Q + \cdots + Q^n) \cdot R & Q^n \end{pmatrix}.$$

Now use the matrix version of the sum of a geometric series.

$$\frac{1}{I - Q} = I + Q + Q^2 + \cdots .$$

So the lower left-hand corner of $P^\infty$ is $N \cdot R$. $Q$ is a matrix of nonnegative numbers less than 1. If $q$ is the largest entry in $Q$ and $q_i$ is the sum of the $i$th row then the $(i, j)$ entry of $Q^n$ is bounded by $q_i^n q$. Thus $Q^n$ tends to the 0 matrix,

$$P^\infty = \begin{pmatrix} I & 0 \\ N \cdot R & 0 \end{pmatrix}$$

and we can read off the result. □

Given a Markov process $P$, the main goal in a finite mathematics course is to determine if the process is asymptotically stable and if so to find the eventual state distribution. Ostensibly, this is a problem in analysis. However, it can be solved by discrete means. Determine (by repeated multiplication) if the matrix $P$ (or in the reducible case the $R$ of Theorem III) is regular. If it is not the process is not asymptotically stable. If it is, then find the stochastic solution to $\bar{x}P = \bar{x}$. Regular matrices are a tool but not the goal of this procedure. For regular Markov processes these arguments justify a 'discrete' solution of a 'continuous' problem. One can find an asymptotically stable vector by checking whether the matrix is regular and finding its fixed point algebraically.

We have given proofs that are accessible to undergraduates of the fundamental theorem of finite Markov processes. There are several more general contexts for these results. First the restriction that the matrices be stochastic can be weakened to demanding only that the entries be nonnegative. The Perron-Frobenius theory (e.g., [1], [7]) uses methods of elementary linear algebra and simple estimates to investigate the characteristic values of a nonnegative matrix. While this development provides a generalization of the results it still fails to answer an important question.

What crucial property of the situation discussed here 'causes' the existence of (asymptotically stable) fixed points? In fact, there are several possible explanations. Thus, a topologist can note that the set of $n$-ary probability vectors is homeomorphic to the $(n - 1)$-disk and conclude Theorem I from the Brouwer fixed point theorem. Alternatively, the linearity and continuity of the function and the fact that the domain is a closed convex set can be exploited as in **[6]** to deduce Theorem I from the Markov-Kakutani fixed point theorem.

Both of these explanations cover only the existence of a fixed point, not the asymptotic stability. The general theory of Lipschitz operators (e.g., **[4]**) provides a context to discuss both phenomena. Let $C$ be a subset of a Banach space $X$. A function $f$: $C \to X$ is said to be of Lipschitz class $K$ for a positive real $K$ if for all $x, y \in C$, $\|f(x) - f(y)\| \leqslant K\|x - y\|$. If $K < 1$ then $f$ is called contractive and if $K = 1$ then $f$ is called nonexpansive. A contractive map is asymptotically stable; moreover, under appropriate conditions on $X$ and $C$ a nonexpansive map has a fixed point. (In particular it suffices for $X$ to be a Hilbert space and $C$ to be closed, convex, and bounded.) Thus the distinction between Theorems I and II can be precisely reflected in a general setting.

REFERENCES


1. D. R. Cox and H. D. Miller, The Theory of Stochastic Processes, Wiley, New York, 1968.
2. Joel Franklin, Mathematical methods in economics, *Amer. Math. Monthly*, 90 (1983) 229–244.
3. M. Girault, Stochastic Processes, Springer-Verlag, Heidelberg, 1966.
4. V. I. Istrăţescu, Fixed Point Theory, D. Reidel Publishing Co., Boston, 1981.
5. J. G. Kemeny and J. L. Snell, Finite Markov Chains, Van Nostrand, New York, 1960.
6. George Maltese, A simple proof of the fundamental theorem of Markov chains, *Amer. Math. Monthly*, 93 (1985) 629–630.
7. E. Seneta, Non-negative Matrices and Markov Chains, Springer-Verlag, Heidelberg, 1980.

# PROBLEMS AND SOLUTIONS



*For instructions about submitting **proposed** problems for publication in this department see the inside front cover. Please include solutions, relevant references, etc.*

*An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.*

***Solutions** should be sent to the address given on the inside front cover. Two copies suffice.*

*A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a* MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

*For instructions about submitting solutions of Problems, which should be mailed before June 30, 1989, see the inside front cover. Please place the solver's name and mailing address on each (double-spaced) sheet. Include a self-addressed card or label if an acknowledgement is desired.*

## ELEMENTARY PROBLEMS

E 3307. *Proposed by Peter Andrews, Wilfrid Laurier University, Waterloo, Ontario, Murray Klamkin, University of Alberta, Edmonton, Alberta, and Edward T. H. Wang, Wilfrid Laurier University.*

The celebrated Morley triangle of a given triangle $ABC$ is the equilateral triangle whose vertices are the intersections of adjacent pairs of internal angle trisectors of $ABC$. If $s$, $R$, $r$, $F$ and $s_M$, $R_M$, $r_M$, $F_M$ are the semiperimeter, the circumradius, the inradius, and the area, respectively, of $ABC$ and its Morley triangle, determine the maximum of (i) $s_M/s$, (ii) $R_M/R$, (iii) $r_M/r$, and (iv) $F_M/F$.

E 3308. *Proposed by Ion Cucurezeanu, Liceul 10 Constanta, Romania.*

Suppose $n$ is a positive integer greater than 4. Prove that the following two conditions are equivalent:
(i) both $n$ and $n+1$ are composite,
(ii) the integer nearest to $(n-1)!/(n^2+n)$ is even.

E 3309. *Proposed by Donald E. Knuth, Stanford University, Stanford, CA.*

Suppose $a$, $b$, $c$ are nonnegative integers with $b > a$. If $0 < \theta < 1$ and $0 \leqslant x \leqslant 1$, prove that

$$(-1)^{b-a-1} \sum_{k=0}^{c} \binom{c}{k} \binom{a+\theta}{b+k} x^k > 0.$$

Here (as usual)

$$\binom{t}{k} = t(t-1)\cdots(t-k+1)/k!$$

for any real $t$ and positive integer $k$, while $\binom{t}{0} = 1$.

E 3310. *Proposed by Albert Wilansky, Lehigh University, Bethlehem, PA.*

I have a secret positive integer $u$, not exceeding some predetermined bound $N$. If you give me any positive integer $n$, I will tell you whether or not $u + n$ is prime.

(a) Give a number $G(N)$ and a procedure such that you can always determine the value of $u$ in at most $G(N)$ such trials.

(b) Give a number $F(N)$ such that no strategy will always determine $u$ in fewer than $F(N)$ such trials.

E 3311. *Proposed by Sydney Bulman-Fleming and Kenneth McDowell, Wilfrid Laurier University, Waterloo, Ontario.*

Suppose $S$ is a monoid containing elements $a$ and $b$ such that every element of $S \times S$ is of the form $(au, bu)$ for some $u$ in $S$ (i.e., $S \times S$, considered as a right $S$-set, is cyclic).

(a) Show that $S$ must be a singleton if $S$ is any one of the following: finite, commutative, idempotent, or inverse.

(b) Show that $S$ need not be a singleton in general.

E 3312 . *Proposed by Robert Blakley and Douglas Hensley, Texas A & M University, College Station, Texas.*

Given positive integers $m, n$ with $n/2 < m < n$, suppose $n$ tennis matches are to be scheduled on $m$ successive days, with at least one match per day. Determine the largest positive integer $K$ such that, no matter how the matches are scheduled, for each $k \leqslant K$ there is a consecutive sequence of days on which a total of exactly $k$ matches are scheduled.

## SOLUTIONS OF ELEMENTARY PROBLEMS

### An Irreducible Polynomial

E 3008 [1983, 482]. *Proposed by Roger Cuculière, Paris.*

Show that for $n \geqslant 2$ the polynomial $x^n - x^{n-1} - x^{n-2} - \cdots - x - 1$ is irreducible over the rationals.

*Solution I by F. Dodd, University of South Alabama, Mobile, and N. J. Lord, Tonbridge School, Kent, England, independently.* Let $P_n(x) = x^n - x^{n-1} - x^{n-2} - \cdots - x - 1$ for each $n \geq 2$ and consider $Q_n(x) = (x-1)P_n(x) = x^{n+1} - 2x^n + 1$. Now $Q_n(1) = 0$ and $Q_n'(r) < 0$ for $1 < r < 2n/(n+1)$, so that $Q_n(r) < 0$ for every $r$ satisfying $1 < r < 2n/(n+1)$. If $|z| = r$, $1 < r < 2n/(n+1)$, we have $|1 - 2z^n| \geq 2|z|^n - 1 = 2r^n - 1 > r^{n+1} = |z|^{n+1}$, which shows that the function $1 - 2z^n$ (strictly) dominates the function $z^{n+1}$ on the circle $|z| = r$. Thus Rouché's theorem implies that $Q_n(z)$ and $1 - 2z^n$ have exactly the same number of zeros

inside $|z| = r$, for every $r$ satisfying $1 < r < 2n/(n+1)$, and hence the same number of zeros inside or on $|z| = 1$. The function $1 - 2z^n$ obviously has $n$ zeros inside $|z| = 1$, while $Q_n(z) = 0$ with $|z| = 1$ forces $z = 1$. Thus $Q_n(z)$ has exactly $(n-1)$ zeros satisfying $|z| < 1$. Since $P_n(1) < 0 < P_n(2)$, we see that $P_n(z)$ has $(n-1)$ zeros satisfying $|z| < 1$ and one zero satisfying $|z| > 1$.

If $P_n(x)$ were reducible over the rationals, Gauss's Lemma would tell us that $P_n(x) = G(x)H(x)$ for suitable monic polynomials $G(x)$, $H(x)$ of positive degrees with coefficients which are rational integers. One of these polynomials, say $G(x)$, must have all its zeros of modulus strictly less than one. However this implies that $|G(0)| < 1$, contradicting the fact that the constant term of the polynomial $G(x)$ must be a nonzero integer.

*Solution II by L. E. Mattics, University of South Alabama, Mobile.* Denote the given polynomial by $P_n(x)$ and $(x-1)P_n(x)$ by $Q_n(x)$. The proposition is easily checked by the rational root theorem for $n = 2, 3$, so from now on we assume $n \geq 4$. By Descartes' rule of signs, $P_n(x)$ has precisely one positive real root $r$ and it is easily checked that $P_n(\sqrt{3}) < 0$, so that $r > \sqrt{3}$.

If $P_n(s) = 0$ and $|s| > 1$, then $Q_n(s) = s^{n+1} - 2s^n + 1 = 0$ and so $|s|^n|s-2| = 1$. We have $2 \leq |s-2| + |s| = |s|^{-n} + |s|$, so that $Q_n(|s|) \geq 0$. Since $Q_n(x) < 0$ for $1 < x < r$, we must have $|s| \geqslant r$. On the other hand, if $P_n(t) = 0$ and $|t| < 1$ then $1 = |t-2|\,|t|^n \leq 3|t|^n$. It follows that the absolute value of the product of all roots $t$ of $P_n(x)$ with $|t| < 1$ is at least $1/3$, so that $r$ is the only root of $P_n(x)$ whose absolute value exceeds 1. (If there were two such roots, then the absolute value of their product would be at least $r^2 > 3$ and then $|P_n(0)| > 1$; this argument also shows that $r$ is not a multiple root of $P_n(x)$.)

Now suppose that $G(x)$ and $H(x)$ are monic polynomials with integer coefficients such that $P_n(x) = G(x)H(x)$ and $G(r) = 0$. Then if $H(x)$ had positive degree, its roots would have absolute value less than 1 and so $|H(0)| < 1$. But the constant term of $H(x)$ must be $\pm 1$. We are done.

*Editorial comment.* M. J. DeLeon noted that the result of the problem is contained in the following theorem of Alfred Brauer: If $a_1, a_2, \ldots, a_n$ are integers with $a_1 \geq a_2 \geq \cdots \geq a_n > 0$, then the polynomial $x^n - a_1x^{n-1} - a_2x^{n-2} - \cdots - a_n$ is irreducible over the rationals. Cf. Alfred Brauer, On algebraic equations with all but one root in the interior of the unit circle, *Math. Nachrichten,* 4 (1950/51) 250–257.

The positive root of the polynomial considered in this problem (or more generally of the polynomial considered in Brauer's Theorem) is an example of what is known as a PV-number, which is a real algebraic integer greater than 1 all of whose conjugates lie inside the unit circle. Cf. J. W. S. Cassels, *An Introduction to Diophantine Approximation,* Cambridge University Press, 1957, Chapter VIII.

Solved also by R. Breusch, I. Connell, A. A. Jagers (The Netherlands), K.-W. Lau (Hong Kong), O. P. Lossers (The Netherlands), H. Niederreiter (Austria), and the proposer.

**A Complex Cyclic Sequence**

E 3181 [1986, 812]. *Proposed by Zalman Rubinstein, University of Colorado, Boulder.*

Let $f(z) = z^2 + z$. Construct a cyclic sequence for $f(z)$, that is, a sequence $\{z_n\}_{-\infty}^{\infty}$ of nonzero complex numbers with $z_{n+1} = f(z_n)$ for all integers $n$ and such

that

$$\lim_{n\to\infty} z_n = \lim_{n\to -\infty} z_n = 0.$$

*Solution by O. P. Lossers, Eindhoven University of Technology, The Netherlands.* For a function $h:\mathbb{C}\to\mathbb{C}$, we let $h^{(n)}$ denote the $n$th iterate of $h$, for each natural number $n$. Consider the two functions $f, g:\mathbb{C}\to\mathbb{C}$ defined by $f(z) = z^2 + z$ and $g(z) = \sqrt{z + (1/4)} - \frac{1}{2}$, where $\sqrt{re^{i\theta}} = \sqrt{r}\,e^{(1/2)i\theta}$ with arguments expressed using $r \geq 0$ and $-\pi < \theta \leq \pi$. Note that $g$ is a right inverse of $f$. We solve the problem by exhibiting a $z_0 \in \mathbb{C} - \{0\}$ such that $f^{(n)}(z_0) \neq 0$, $g^{(n)}(z_0) \neq 0$ for all $n \in \mathbb{N}$ and $f^{(n)}(z_0) \to 0$, $g^{(n)}(z_0) \to 0$.

Let $L$ be the set of negative real numbers with magnitude less than 1. A simple computation yields $f(L) \subseteq L$ and $f^{(n)}(z) \to 0$ for all $z \in L$. We will choose a suitable $z_0 \in L$.

Let $A$ be the set of complex numbers having positive real and imaginary parts. For each $z \in A$, we claim (i) $g(z) \in A$, (ii) $|g(z)| < |z|$, and (iii) $g^{(n)}(z) \to 0$. Here (i) can be seen by writing $\sqrt{z + (1/4)} = \sqrt{x + iy + (1/4)} = a + ib$ and squaring both sides to show that $a > \frac{1}{2}$ and $b > 0$ and (ii) can be seen by rewriting $g(z)$ as $g(z) = z/(\sqrt{z + (1/4)} + \frac{1}{2})$ and noting that $|\sqrt{z + (1/4)} + \frac{1}{2}| > \operatorname{Re}\sqrt{z + (1/4)} + \frac{1}{2} > 1$. To show (iii), take any $z \in A$ and write $g^{(n)}(z) = z_n = r_n e^{i\theta_n}$, where $r_n > 0$ and $0 < \theta_n < \pi/2$. By (ii), the magnitudes converge, say $r_n \to r \geq 0$. If $r > 0$, then $r_{n-1} = |f(z_n)| = r_n \cdot |r_n e^{i\theta_n} + 1|$ implies $|r_n e^{i\theta_n} + 1| \to 1$, or equivalently that $r_n(r_n + 2\cos\theta_n) \to 0$, which is impossible. Hence $r = 0$ and (iii) holds.

It suffices to select $z_0 \in L$ such that $g^{(n)}(z_0) \in A$ for some $n \in \mathbb{N}$ (which implies that $g^{(n)}(z_0) \neq 0$ for all $n \in \mathbb{N}$). The number $z_0 = -13/16$ will do the job, since

$$g(z_0) = -\frac{1}{2} + \frac{3}{4}i$$

and

$$g^{(2)}(z_0) = \frac{1}{2}\left[\sqrt{\frac{\sqrt{10} - 1}{2}} + i\sqrt{\frac{\sqrt{10} + 1}{2}}\right] - \frac{1}{2} \in A.$$

Also solved by A. Zulauf (New Zealand), University of South Alabama Problem Group, and the proposer. One incorrect solution was received.

### New Polygons from Old

E 3183* [1987, 71]. *Proposed by Murray Klamkin, University of Alberta, Edmonton, Canada.*

Let $P'$ denote the convex $n$-gon whose vertices are the midpoints of the sides of a convex $n$-gon $P$. Determine the extreme values of

(i) Area $P'$/Area $P$,

(ii) Perimeter $P'$/Perimeter $P$.

*Solution by David B. Secrest (student), University of Illinois, Urbana.* The extreme values are summarized in the following table:

| $n$ | 3 | 4 | 5 | $\geq 6$ |
|---|---|---|---|---|
| $\max\left(\dfrac{\text{area } P'}{\text{area } P}\right)$ | $\dfrac{1}{4}$ | $\dfrac{1}{2}$ | $\dfrac{3}{4}$ | 1 |
| $\min\left(\dfrac{\text{area } P'}{\text{area } P}\right)$ | $\dfrac{1}{4}$ | $\dfrac{1}{2}$ | $\dfrac{1}{2}$ | $\dfrac{1}{2}$ |
| $\max\left(\dfrac{\text{perim } P'}{\text{perim } P}\right)$ | $\dfrac{1}{2}$ | 1 | 1 | 1 |
| $\min\left(\dfrac{\text{perim } P'}{\text{perim } P}\right)$ | $\dfrac{1}{2}$ | $\dfrac{1}{2}$ | $\dfrac{1}{2}$ | $\dfrac{1}{2}$ |

We divide the proof into seven cases.

*Case* 1. $n = 3$. The triangle $P'$ is similar to $P$ with sides exactly half as long; the conclusion is immediate for all four extrema.

*Case* 2. $n = 4$, *Area Ratio.* Given a quadrilateral $E_1E_2E_3E_4$, let $E_1'$, $E_2'$, $E_3'$, and $E_4'$ be the midpoints of the sides $\overline{E_1E_2}$, $\overline{E_2E_3}$, $\overline{E_3E_4}$, and $\overline{E_4E_1}$, respectively. By similar triangles, $\text{area}(E_1'E_2E_2') = \frac{1}{4}\,\text{area}(E_1E_2E_3)$ and

$$\text{area}(E_3'E_4E_4') = \frac{1}{4}\,\text{area}(E_3E_4E_1),$$

so

$$\text{area}(E_1'E_2E_2') + \text{area}(E_3'E_4E_4') = \frac{1}{4}\,\text{area}(E_1E_2E_3E_4).$$

Likewise,

$$\text{area}(E_4'E_1E_1') + \text{area}(E_2'E_3E_3') = \frac{1}{4}\,\text{area}(E_1E_2E_3E_4),$$

so

$$\frac{\text{area}(E_1'E_2'E_3'E_4')}{\text{area}(E_1E_2E_3E_4)} = \frac{1}{2}.$$

*Case* 3. $n \geq 6$, *Maximum Area Ratio.* A degenerate $n$-gon,

$$P = E_1E_2 \cdots E_n, \quad \text{with } E_1 = E_2, \quad E_3 = E_4, \quad \text{and } E_5 = E_6 = \cdots = E_n$$

will have area $P'/\text{area } P = 1$, which is the largest the ratio can be if $P$ is convex.

*Case* 4. $n \geqslant 5$, *Minimum Area Ratio.* Given an $n$-gon, $P = E_1E_2 \cdots E_n$, draw all the "corner triangles", i.e. the ones made by connecting three consecutive vertices of the $n$-gon. Let $X$ be some point inside corner triangle $E_2E_3E_4$. $X$ cannot be inside any of the other corner triangles except possibly $E_1E_2E_3$ and $E_3E_4E_5$ because all the others are disjoint from $E_2E_3E_4$. However, $E_1E_2E_3$ and $E_3E_4E_5$ are disjoint, so $X$ can be inside at most one of them. In other words, no point inside the $n$-gon $P$ can be inside more than two of the corner triangles. Thus

$$\sum \text{area(corner triangles)} \leq 2\,\text{area } P.$$

But the triangles made by connecting midpoints of adjacent sides of $P$ (which will be referred to hereafter as the "midpoint corner triangles") each has $1/4$ the area of the corresponding corner triangle, so that

$$\sum \text{area(midpoint corner triangles)} \leq \frac{1}{2} \text{ area } P.$$

Adding area $P'$ to both sides of this inequality and noticing that $P'$ and the midpoint corner triangles together make up $P$, we get

$$\text{area } P \leq \frac{1}{2} \text{ area } P + \text{area } P'$$

or

$$\frac{\text{area } P'}{\text{area } P} \geqslant \frac{1}{2}.$$

It is possible to attain this minimum ratio of $\frac{1}{2}$ by setting $E_3 = E_4 = \cdots = E_n$.

*Case* 5. $n = 5$, *Maximum Area Ratio.* Given a convex pentagon, $P = E_1E_2E_3E_4E_5$, if we could prove that

$$\text{area}(E_1E_3E_5) \leq \text{area}(E_5E_1E_2) + \text{area}(E_2E_3E_4) + \text{area}(E_4E_5E_1),$$

then, by adding area$(E_1E_2E_3)$ + area$(E_3E_4E_5)$ to both sides, we could get

$$\text{area } P \leq \sum \text{area(corner triangles)}$$

or

$$\frac{1}{4} \text{ area } P \leq \sum \text{area(midpoint corner triangles)}$$

and it would follow, after adding area $P'$ to both sides, that

$$\frac{1}{4} \text{ area } P + \text{area } P' \leq \text{area } P \quad \text{or} \quad \text{area } P'/\text{area } P \leq \frac{3}{4}.$$

We will show that by choosing a suitable labelling of the vertices of $P$, we can prove the stronger result, that

$$\text{area}(E_1E_3E_5) \leq \text{area}(E_5E_1E_2) + \text{area}(E_2E_3E_4). \tag{1}$$

In the convex pentagon, $P$, there must be a pair of adjacent angles that add to more than $\pi$ since the average sum of pairs of adjacent angles in a pentagon is $6\pi/5$. Assume $\angle E_4 + \angle E_5 > \pi$. (see FIGURE 1). This implies that the extension of $\overline{E_1E_5}$ beyond $E_5$ and the extension of $\overline{E_3E_4}$ beyond $E_4$ intersect. Draw the lines $\ell_1$ and $\ell_2$ through $E_2$ and parallel to $\overline{E_1E_5}$ and $\overline{E_3E_4}$, respectively. Draw $\overline{E_1E_3}$. Pick a point, $D$, on $\overline{E_1E_3}$ so that it is inside the parallelogram formed by $\ell_1$, $\ell_2$, and the extensions of $\overline{E_1E_5}$ and $\overline{E_3E_4}$. Draw lines through $E_4$ and through $E_5$ parallel to $\overline{E_1E_3}$. By performing a reflection of the picture if necessary, we can assume $E_5$ is closer to $\overline{E_1E_3}$ than $E_4$ is.

We establish some inequalities:

$$\text{area}(DE_3E_5) \leq \text{area}(DE_3E_4) \tag{2}$$

because they share a common base, $\overline{DE_3}$, but $E_4$ is farther away from that base.

Likewise, for the base $\overline{E_5E_1}$,

$$\operatorname{area}(E_5E_1D) \leq \operatorname{area}(E_5E_1E_2), \tag{3}$$

and, for the base $\overline{E_3E_4}$

$$\operatorname{area}(DE_3E_4) \leq \operatorname{area}(E_2E_3E_4). \tag{4}$$

Now (2) and (4) yield area$(DE_3E_5) \leq$ area$(E_2E_3E_4)$. By adding (3) to the last inequality we get (1) and thus area $P'$/area $P \leq \frac{3}{4}$.

It is possible to attain this maximum ratio by setting $E_1 = E_2$ and $E_3 = E_4$.

*Case* 6. $n \geq 4$, *Maximum Perimeter Ratio.* An $n$-gon, $P = E_1E_2 \cdots E_n$, which has $E_1 = E_2$ and $E_3 = E_4 = \cdots = E_n$ will have perim $P$ = perim $P'$.

It is impossible to get a ratio larger than one since, by the triangle inequality, the length of one of the sides of $P'$ is at most that of the two half sides of $P$ that it replaces. (This argument works even if $P$ is not convex.)

*Case* 7. $n \geq 4$, *Minimum Perimeter Ratio.* Given an $n$-gon, $P = E_1E_2 \cdots E_n$, let $D_1$ be the point of intersection of $\overline{E_nE_2}$ and $\overline{E_1E_3}$, and in general, $D_k$ the point of intersection of $\overline{E_{k-1}E_{k+1}}$ and $\overline{E_kE_{k+2}}$, where the subscripts are taken mod $n$. (See FIGURE 2). By the triangle inequality,

$$\begin{gathered} E_1E_2 \leq D_1E_2 + E_1D_1 \\ E_2E_3 \leq D_2E_3 + E_2D_2 \\ \vdots \\ E_nE_1 \leq D_nE_1 + E_nD_n. \end{gathered}$$

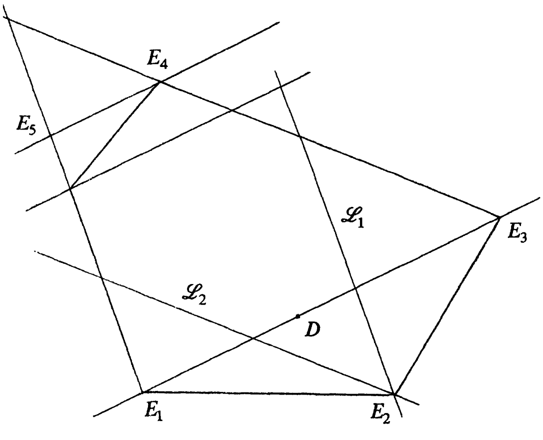



FIG. 1.

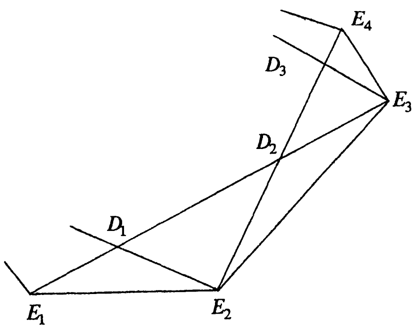



FIG. 2.

Also, $0 \leq D_1D_2 + D_2D_3 + \cdots + D_nD_1$. Adding, we obtain

$$\begin{aligned} \operatorname{perim} P &\leq (D_1E_2 + E_1D_1) + D_1D_2 + (D_2E_3 + E_2D_2) + D_2D_3 + \cdots \\ &= D_1E_2 + (E_1D_1 + D_1D_2 + D_2E_3) + (E_2D_2 + D_2D_3 + D_3E_4) + \cdots \\ &= E_1E_3 + E_2E_4 + \cdots + E_nE_2 \\ &= 2 \operatorname{perim} P', \end{aligned}$$

so

$$\text{perim } P'/\text{perim } P \geq \frac{1}{2}.$$

If $E_1E_2,\ldots, E_n$ is a polygon with $E_2 = E_3 = \cdots = E_n$, then perim $P' = 1/2$ perim $P$.

Solved also by David Callan and jointly by G. Arnold, V. Konečný, and R. Shepler. Partial solution by P. L. Hon (Hong Kong).

### Functions with the Darboux Property

E 3191 [1987, 181]. *Proposed by Gheorghe Răuțu, Centre of Mathematical Statistics, Bucharest, Romania.*

Find all functions $f$: $\mathbb{R} \to \mathbb{R}$ which have the Darboux property and which satisfy the equality

$$f(x+y) = f(x+f(y))$$

for $x$, $y \in \mathbb{R}$.

*Solution by Adam Riese, Wright State University, Dayton, OH.*

We show $f$ must be a constant or the identity function. The range of a Darboux function is connected, and thus is an interval. Let $a$ and $b$ be the infimum and supremum of the range of $F$. With $x = 0$, the condition of the problem yields $f(y) = f(f(y))$. This implies $x = f(x)$ for $x \in (a, b)$.

First suppose $a$ is finite and $f$ is not a constant. Since $f(x) = x$ for $x \in (a, b)$ and $f$ is Darboux, $f(a) = a$. If $f$ is not a constant, then $f(x) = x$ on an interval $[a, a + 2d]$ for some $d > 0$. In this case, any value of $t$ with $0 < t < d$ satisfies $f(a - t) = a + s$ for some $s$ with $s > d$ ($s \leq d$ leads to the contradiction $a = f(t + a - t) = f(t + a + s) = a + s + t$). However, $f(a - t) > a + d$ for all $t$ in $(0, a)$ and $f(a) = a$ contradict the Darboux property. Hence $a = -\infty$ if $f$ is not a constant, and similarly $b = \infty$, and hence $f$ is the identity.

*Editorial Comment.* The term "Darboux property" should have been defined in the statement of the problem. It is the intermediate value property, meaning that if $f(x) = A$ and $f(y) = B$, then on the interval between $x$ and $y$ the function $f$ takes every value between $A$ and $B$. V. C. Williams and E. A. Enneking noted that $f(x) = x - \lfloor x \rfloor$ is a solution to the functional equation but does not satisfy the Darboux property.

Solved also by S. F. Barger, R. B. Eggleton, O. P. Lossers, M. J. Reed, jointly by V. C. Williams & E. A. Enneking, and the proposer. One incorrect solution was received.

### Maximal Sets of Dispersed Vectors

E 3206 [1987, 373]. *Proposed by Chico Problem Group, California State University, Chico.*

Let $S = \{x_1, x_2,\ldots, x_n\}$ be a set of $n$ nonzero vectors in $\mathbb{R}^d$ such that $x_i \cdot x_j \leq 0$ whenever $1 \leq i < j \leq n$.

(a) Prove that $n \leq 2d$, and characterize those $S$ for which $n = 2d$.

(b) What if $x_i \cdot x_j \leq 0$ is replaced by $x_i \cdot x_j < 0$?

*Solution by David M. Wells, Pennsylvania State University, New Kensington.*

We will verify two assertions simultaneously by induction on $d$:

1) $n \le 2d$, with equality if and only if half of the vectors $S$ form an orthogonal basis for $\mathbb{R}^d$ and the rest are negative multiples of these.

2) If $x_i \cdot x_j < 0$ for $i \neq j$, then $n \le d + 1$.

Both assertions are true for $d = 1$. Now suppose $d > 1$, and let $S = \{x_1, \dots, x_n\}$ be a largest set of the desired type. Choose an orthogonal basis for $\mathbb{R}^d$ in which the coordinates of $x_n$ are $(-1, 0, \dots, 0)$. Let $a_i$ be the first coordinate of $x_i$; note $a_i \ge 0$ for $i < n$. Let $P = \{y \in \mathbb{R}^d\colon y \cdot x_n = 0\}$, and let $y_i = \operatorname{Proj}_P x_i$. We can now verify both assertions.

1) $y_i \cdot y_j = x_i \cdot x_j - a_i a_j \le x_i \cdot x_j \le 0$, so by the induction hypothesis there are at most $2d - 2$ nonzero vectors in $S' = \{y_1, \dots, y_{n-1}\}$. Thus $n \le 2d$, with equality only if $S'$ has $2d - 2$ nonzero vectors and $y_i = 0$ for some $i < n$. In this case $x_i$ is a negative multiple of $x_n$, so $x_i$ is orthogonal to $P$, and all vectors in $S$ except $x_i$ and $x_n$ lie in $P$.

2) If $x_i \cdot x_j < 0$ for $i \neq j$, then $y_i \cdot y_j < 0$ also. By the induction hypothesis, there are at most $d$ vectors in $\{y_1, \dots, y_{n-1}\}$, so $n \le d + 1$.

*Editorial comment.* Several readers supplied examples achieving the bound in part (b). For example, Ilan Kozma suggests $x_{d+1} = (-1, \dots, -1)$ and $x_i = (-1, \dots, -1, i, 0, \dots, 0)$ for $i \le n$, where the positive coordinate is the $i$th coordinate. Then $x_i \cdot x_j = -1$ for $i \neq j$.

Also solved by C. Blankespoor, J. Ferrer (Spain), T. Jager, L. R. King, I. Kozma (Israel), O. P. Lossers (The Netherlands), N. Martin (part a only), A. Pedersen (Denmark), J. H. Steelman, Y. L. Wong, Univ. of South Alabama Problem Group, Northridge Problem Solving Seminar, and the proposers.

## A Reluctant Random Walk

E 3213 [1987, 548]. *Proposed by D. M. Bloom, Brooklyn College, CUNY.*

As in Problem E 2276 [1971, 78; 1972, 90], we consider the "Ehrenfest urn" game in which players $A$ and $B$ have between them $n$ cards labeled $1, 2, \dots, n$. At each move, one of the numbers $1, 2, \dots, n$ is chosen at random and the player who has the card with that number must give it to the other player. The game continues until one player has all the cards. Prove that the expected length of the game is

(a) $2^{n-1} - 1$ if $A$ initially has exactly one card.

(b) $2^{n-1} - \frac{1}{2}(n + 1)$ if $A$ is equally likely to start with any number of cards in $\{1, 2, \dots, n\}$.

*Solution by Bruce R. Johnson, University of Victoria, B.C., Canada.*

Let $e_j$ be the expected length of the game, given that player $A$ starts with $j$ cards; we derive a recurrence relation for $e_j$. Conditioning on the outcome of the first move yields

$$e_j = 1 + \frac{n-j}{n} e_{j+1} + \frac{j}{n} e_{j-1},$$

which we rewrite as

$$e_j - e_{j-1} = \frac{n}{j} + \frac{n-j}{j}(e_{j+1} - e_j) \quad \text{for } 1 \le j \le n-1. \tag{1}$$

Applying this recursively yields

$$e_1 - e_0 = \left(\sum_{i=1}^{j}\binom{n}{i}\right) + \binom{n-1}{j}(e_{j+1} - e_j) \quad \text{for } 1 \le j \le n-1.$$

In this expression at $j = n - 1$, use $e_0 = e_n = 0$, $e_{n-1} = e_1$, and the binomial formula to simplify this to $e_1 = 2^{n-1} - 1$, proving (a).

To prove (b), multiply both sides of (1) by $j$ and sum over $j$ to obtain

$$\sum_{j=1}^{n-1} j(e_j - e_{j-1}) = (n-1)n + \sum_{j=1}^{n-1}(n-j)(e_{j+1} - e_j).$$

Combining consecutive terms in the sums simplifies this to

$$ne_{n-1} - \sum_{j=0}^{n-1} e_j = (n-1)n - ne_1 + \sum_{j=1}^{n} e_j.$$

Using $e_n = e_0 = 0$ and $e_{n-1} = e_1 = 2^{n-1} - 1$, collecting like terms, and dividing by $2n$ yields $(1/n)\Sigma_{j=1}^{n} e_j = 2^{n-1} - (n+1)/2$, as desired.

This game has a large expected length, because the player losing is the favorite to win the next move, the further behind the stronger the favorite. Also, the expected length does not change much as a function of the initial distribution. For example, if $n = 10$, then $e_1 = e_9 = 511$, $e_2 = e_8 = 566\frac{2}{3}$, $e_3 = e_7 = 579\frac{1}{3}$, $e_4 = e_6 = 583\frac{1}{3}$, and $e_5 = 584\frac{1}{3}$.

Also solved by J. Fitch, O. P. Lossers (The Netherlands), L. E. Mattics, D. Neuenschwander (Switzerland), A. Pedersen (Denmark), H. L. Stubbs, Western Maryland College Problem Group, and the proposer. One partial solution was received.

### How to Gamble If You Must

E 3219 [1987, 680]. *Proposed by Daniel Rawsthorne, Silver Spring, MD.*

A gambler has no money, but the host of the casino generously allows him to play 100 games of the following type. He may either (1) choose to accept one dollar with no risk or (2) choose an integer $n > 1$, whereupon he wins $n$ dollars with probability $2/(n+1)$ or loses one dollar with probability $(n-1)/(n+1)$. (He must have at least one dollar to choose option (2).) What is an optimal strategy for the gambler if he wishes to leave with \$200 or more, and what is his probability of success using that strategy?

*Solution by Carl Schoen, University of Wisconsin, Eau Claire.*

The player should choose $n = 1$ (certain one dollar win) on each odd play and $n = 99 + k$ on play $k$, where $k$ is even. He should continue this until winning on an even play; from this point on he should choose $n = 1$ for each of the remaining plays. His probability of \$200 is then exactly .5.

For any chosen option, the player's expected win is one dollar on a single play. Therefore, after 100 plays, his expected total win $W$ is 100, regardless of strategy. Since $W \ge 0$ always, the probability of $W \ge 200$ therefore cannot exceed .5 for any strategy, and can equal .5 only if $W = 200$ and $W = 0$ are the only possible outcomes.

We show that the strategy described above is the only strategy that achieves this. If $W = 200$ or $W = 0$ are the only outcomes of the strategy, then after 99 plays the total must be 1 or 199, and on the 100th play the player must choose $n = 199$ or $n = 1$, respectively, since any other situation yields positive probability for some other value of $W$. If 1 and 199 are the possible totals after 99 plays, then the possible totals after 98 plays must be 0 or 198, and the player must choose $n = 1$ on the 99th play, for the same reason. In general, if 0 and $100 + k$ are the only possible outcomes after the $k$th play, with $k$ even, then the total before that play must be 1 or $99 + k$, and the player must choose $n = 99 + k$ or $n = 1$, respectively. Similarly, if 1 and $100 + k$ are the only possible outcomes after the $k$th play, with $k$ odd, then the total before that play must be 0 or $99 + k$, with the player choosing $n = 1$ on the $k$th play. Following this backward yields the strategy described above.

Also solved by F. T. Bruss, A. Gorfin, J. W. Grossman, C. Hurd, R. Jacobson, J. Lawrence, O. P. Lossers (The Netherlands), W. A. Newcomb, and the proposer. Partially solved by C. Aschbacher. Three incorrect solutions were received.

### Within ε of Fermat's Last Theorem

E 3223 [1987, 681]. *Proposed by the editors (a modification of a problem proposed by the late William F. Eberlein).*

Given real numbers $\varepsilon > 0$ and $p > 1$, show that for all sufficiently large positive real numbers $K$ the curve $x^p + y^p = K^p$ comes within distance $\varepsilon$ of a point with positive integral coordinates.

*Solution I by Timothy S. Lewis, Eastman Kodak Company, Rochester, NY.* Given $\varepsilon$ and $p$, choose $K$ large enough so that

$$\left[\left(\frac{K}{K-1}\right)^p - 1\right]^{(p-1)/p} < \varepsilon. \tag{*}$$

Let $m = \lfloor K - 1 \rfloor$ and $n = \lfloor L - 1 \rfloor$, where $L = (K^p - m^p)^{1/p}$. The distance from $(m, n)$ to the curve is less than the horizontal distance from $(m, n)$ to the line tangent at $(m, L)$; call this horizontal distance $h$. The slope of the curve at $(m, L)$ is $-(m/L)^{p-1}$, so

$$h = (L-n)(L/m)^{p-1} = (L-n)\left(\frac{K^p}{m^p} - 1\right)^{(p-1)/p}$$

Since $L - n \leq 1$ and $m \geq K - 1$, this quantity is bounded by the left side of (*).

*Solution II by the editors.* As $x$ runs from 0 to $(2pK^{p-1})^{1/p}$, $y$ decreases from $K$ to $K - 2 + O(1/K)$. Furthermore, on this interval

$$\left|\frac{dy}{dx}\right| = \left(\frac{x}{y}\right)^{p-1} < \left(\frac{(2pK^{p-1})^{1/p}}{K-3}\right)^{p-1} = O(K^{(1-p)/p}).$$

Since $y$ varies in steps of at most $O(K^{(1-p)/p})$, it must come within $O(K^{(1-p)/p})$ of $\lfloor K - 1 \rfloor$ for an integral value of $x$.

*Editorial Comment.* Delany located a suitable point near the line $y = x$. Delany and Lossers noted that the result also holds for $0 < p < 1$.

Solved also by J. Delany, O. P. Lossers (The Netherlands), S. Phillip, J. Sturm, and the University of South Alabama Problem Group.

## ADVANCED PROBLEMS

6592. *Proposed by Adolf Hildebrand, University of Illinois at Urbana-Champaign.*

For positive real $k$ let $\sigma_k(n)$ denote the sum of the $k$th powers of the (positive) divisors of the natural number $n$. Show that there exists a number $k_0$ satisfying $1 < k_0 < 2$ such that

(i) if $k > k_0$, then $\{\sigma_k(n+1) - \sigma_k(n)\}(-1)^{n+1} > 0$ for all sufficiently large $n$,

(ii) if $0 < k < k_0$, then $\{\sigma_k(n+1) - \sigma_k(n)\}(-1)^{n+1}$ changes sign infinitely often. (Cf. Problem 6555.)

6593. *Proposed by Michel Balazard, Limoges, France.*

Suppose $E$ is a two-dimensional vector space of continuous real-valued functions on $[0,1]$ such that for every $x \in [0,1]$ there exists $g$ in $E$ with $g(x) \neq 0$.

Suppose $S$ is a nonempty subset of $E$ such that

$$G(x) = \sup_{g \in S} g(x) < +\infty$$

for every $x \in [0,1]$. Show that $G$ is continuous and give an example to show that the result fails if $E$ is three-dimensional instead of two-dimensional.

6594. *Proposed by Stephen M. Gersten, University of Utah, Salt Lake City.*

Let $G$, $A$, and $B$ be groups with $A$ and $B$ both free groups. Let $\phi$: $A \to G * B$ be a homomorphism and let $\alpha$: $G * A \to G * B$ be given by $\alpha = (i, \phi)$, where $i$ is the injection $G \to G * B$; that is

$$(i, \phi): g_1 a_1 \cdots g_n a_n \mapsto g_1 \phi(a_1) \cdots g_n \phi(a_n).$$

Consider a commutative diagram

$$\begin{tikzcd} G * A \arrow[r, "\alpha"] \arrow[d, "p_A"] & G * B \arrow[d, "p_B"] \\ A \arrow[r, "\beta"'] & B \end{tikzcd}$$

where the vertical arrows send $G$ to $\{1\}$ and are the identities on $A$ and $B$, respectively. Prove that if $\beta$ is injective, then $\alpha$ is also injective.

## SOLUTIONS OF ADVANCED PROBLEMS

6540 [1987, 303]. *Proposed by Mo Song-Qing, Institute of Applied Physics and Computational Mathematics, Beijing, China.*

Suppose $x$ is a given real number greater than 1. Let $a_n = \lfloor x^n \rfloor$ for $n = 1, 2, \ldots$, where $\lfloor x \rfloor$ is the integral part of $x$. Let $S$ be the infinite decimal $S = 0.\, a_1 a_2 a_3 \ldots$,

where the notation indicates the expansion formed by writing down the decimal digits of $a_1, a_2, a_3, \ldots$ in turn. (For example, if $x = \pi$, then $S = 0.393197\ldots$.) Is it possible for $S$ to be rational?

*Solution by Michael Filaseta, University of South Carolina, Columbia.* It is impossible for $S$ to be rational. For every positive integer $m$ there is an $\varepsilon > 0$ sufficiently small so that $10^{\varepsilon}$ is in the interval $[1, 1 + 10^{-m}]$. Next, for every such $m$ and $\varepsilon$ and for every $\alpha > 0$ there are infinitely many positive integers $n$ such that $\lfloor n\alpha \rfloor \geqslant m$ and $\{n\alpha\}$, the fractional part of $n\alpha$, is in the interval $[0, \varepsilon)$. For $\alpha = \log_{10} x$, $k = \lfloor n\alpha \rfloor$, and $n$ as above we see that

$$y = x^n = 10^{n\alpha} = 10^{\lfloor n\alpha \rfloor} 10^{\{n\alpha\}} < 10^k 10^{\varepsilon}$$

satisfies

$$10^k \leqslant y < 10^k + 10^{k-m}.$$

Hence $\lfloor x^n \rfloor$ begins with the digit 1 followed by at least $m$ zeros. Hence the decimal expansion of $S$ has arbitrarily long blocks of zeros, a conclusion slightly stronger than irrationality.

*Editorial remark.* All solutions received were essentially the same as Filaseta's. Filaseta also provided some other related results. For example, let $f$ be an integer valued arithmetic function such that as $n \to \infty$ we have $\limsup f(n) = \infty$, but $f(n+1)/f(n) < \sqrt{10}$ for $n$ sufficiently large. Then $0.f(1)f(2)f(3)\ldots$ is irrational. He also mentioned some further analogues due to Carl Pomerance.

Special cases of the result of this problem were given by various authors in the *Journal of Number Theory*, 13 (1981) 268–269, 19 (1984) 248–253, 25 (1987) 211–212.

Solved also by Neal Felsinger, Bruce Reznick, and The University of South Alabama Problem Group.

# REVIEWS

EDITED BY JOSEPH KONHAUSER
*Department of Mathematics, Macalester College, St. Paul, MN 55105*

*Littlewood's Miscellany*. Edited by Béla Bollobás. Cambridge University Press, Cambridge, 1986, viii + 200 pp.

RALPH P. BOAS
*Department of Mathematics, Northwestern University, Evanston, IL 60201*

Unless you are not only a mathematician, but also either over 50 years old, or a specialist in certain branches of classical analysis, you may never even have heard of J. E. Littlewood. (Not long ago, he was mentioned in an article in *Science*, but that article mistakenly gave him the wrong initials. I wrote in to correct the error, and was told that nobody else had noticed it.) However, up to about 1940, classical analysis was dominated by three mathematicians, who were, as one of my teachers put it, G. H. Hardy, J. E. Littlewood, and "Hardy and Littlewood." There is a story (not included in this book) that some 75 years ago Landau thought that Littlewood was a pen name for Hardy, who (in this interpretation) wanted not to seem to be dominating British analysis. (Indeed, for many years Hardy had a paper in every volume of the *Proceedings of the London Mathematical Society*.) In search of an existence proof, Landau visited Cambridge, saw a great deal of Hardy, nothing of Littlewood, and returned to Göttingen convinced of the truth of his conjecture. Of course, Littlewood was real, and was (even without counting his joint work with Hardy) one of the most important analysts of his day.

The popular stereotype of a mathematician may have been influenced by the publicity about Einstein, who was of course a great physicist but an indifferent mathematician. A mathematician is expected to look unworldly, and peculiar in some respect—either retarded or very brilliant, as one novelist put it. However, when I first saw Littlewood in 1938, I thought that he looked more like a teller in a bank than like a mathematician. The new edition of his book has several photographs of him at various ages. In the one of him lecturing (page 14) he seems to have stopped to think; when he did this, he stood absolutely still. Hardy, on the other hand, was constantly in motion. There is no truth—on Hardy's own evidence—that Hardy (in the same situation) ever said, "It's obvious... Is it obvious?", left the room for fifteen minutes, and then returned saying "Yes, it's obvious." You need to understand that "obvious," as used by Hardy and Littlewood, meant something like "It seems plausible, and a moment's thought will suggest a proof." Of course, a "moment" for Hardy or Littlewood would have been much longer for the average listener.

The year when I was in Cambridge Littlewood was lecturing on complex analysis. There was, in those days, no way to find out what he had to say except to attend his lectures. There was no photocopying and none of the modern exchange of preprints. Part of Littlewood's book on complex analysis had existed in galley proof for many years, and copies of it circulated in Cambridge, but it was not published until several years later. There existed a "Guide to the Literature of Mathematics and Physics"; it was about the size of the Combined Membership List, but half its

thickness, and included practically all the books there were; and a reasonably well-informed mathematician would have at least heard of all of them.

Indeed, the world in which Hardy and Littlewood flourished was not only smaller than ours, by a factor of at least 10, but simpler. Hardy could, and always did, submit handwritten manuscripts to the printer. Examinations (many fewer than we are accustomed to) were set in moveable type and printed. Mathematicians enjoyed a much higher status in Great Britain than they do here and now, as the preface to this book brings out (but their salaries were not all commensurate with their status).

Ph.D.s were less common than in the United States, and much less common than they now are. Neither Hardy nor Littlewood had a Ph.D., although by 1938 a demand for the Ph.D. had developed even at Cambridge, and both of them directed Ph.D. candidates. It was explained to me that there was a higher degree, the Sc.D., that you could apply for on the strength of your published work. Since people like Hardy and Littlewood obviously qualified, there was no reason for them to apply unless they had wanted the more spectacular academic robes that went with the Sc.D. Hardy, at least, had little respect for the Ph.D. degree, and was not above writing a thesis for a student. One such student then asked Hardy to give him a letter saying that he had written a good thesis, so that he could get a better job in his own country. This, Hardy could not conscientiously do, but he advised the student to show the thesis to Littlewood; Littlewood saw that it was a good thesis, wrote the letter, and the student got the job. I once repeated this story to a compatriot of the student's, and he said, "Oh, I know that man"; but he wouldn't tell me who it was.

Hardy used to describe how difficult it was for his generation to learn Lebesgue integration. Littlewood was younger, but it is clear from his account of his mathematical education that he too had to learn it after he was a well-established mathematician. They must have studied it in French, and always said "p.p." (presque partout) for "almost everywhere." Littlewood, even in his fifties, was quite receptive of new ideas. One day I was in the mathematics library (which was kept in the quarters of the Cambridge Philosophical Society), glanced at the new journals, and saw nothing promising. A little later, Frank Smithies came in and asked if there were any interesting new journals. "No," I said, "only the *Proceedings of the Lund Physiographical Society*." Frank went over, opened the *Proceedings*, and found Thorin's proof of the Riesz convexity theorem, which created great excitement in Cambridge. The proof so fascinated Littlewood that it is described, in his own inimitable style, in this book.

Littlewood's book was originally published in 1953 as *A Mathematician's Miscellany*. (The first edition was originally turned down by the publisher that now publishes the revised edition.) It is very different from most books on mathematics for a wide audience, partly because not all of it even pretends to be about mathematics. It was intended to be entertaining, and there is something in it to entertain almost everybody. There are some examples of mathematics with minimum raw material; some technically elementary but more difficult mathematics; an account of Littlewood's own mathematical education; and a delightful essay on large numbers. Some of the chapters are full of gossip (not all of which is intelligible in another place and time); some are full of witty remarks (not all by J. E. L.).

One essay is on the discovery of the planet Neptune. Here Littlewood describes how *he* would have gone about discovering Neptune (more simply than Adams did).

This may seem surprising, but less so if you realize that the undergraduate curriculum in Littlewood's Cambridge included a great deal of applied mathematics and mathematical physics. This was still true three decades later. My friend Frank Smithies did his undergraduate work at Edinburgh (which had much the same kind of curriculum) at about the same time that I was at Harvard, but when we met I discovered that he was about two years ahead of me: he knew everything that I did, plus all that applied mathematics.

Another essay speculates about how Newton discovered the facts about the gravitational attraction of a sphere. There are also some notes on ballistics, and a discursus on how not to write mathematics—the last includes a devastating parody of a style that has unfortunately not yet completely disappeared. The editor has added a substantial amount of material that Littlewood amassed after 1953, many photographs, and a very informative biographical preface. (On p. 6, the reference to p. 100 should be to p. 110.)

*Forever Undecided: A Puzzle Guide to Gödel.* By Raymond Smullyan. Alfred A. Knopf, Inc., 1987, xii + 257 pp.

Craig Smoryński
*Mathematical Institute, State University of Utrecht, The Netherlands*

A few years ago the *Annals of Mathematical Logic* underwent a name change: it is now the *Annals of Pure and Applied Logic.* I, for one, do not like the new name. It follows, does it not, that if there are pure and applied logics there must be pure and applied logicians, just as there are pure and applied mathematicians? And will it not also follow, as it has in mathematics, that these two groups will not speak to each other? I beg all readers to swamp the APAL (an appalling acronym) editorial board with mail asking to change the name back and preserve the status quo. For, as I hope to illustrate below, the difference between pure and applied logic is merely one of where the application lies—in the sacred world of mathematics or the profane world of everyday affairs. The two types of applications use the same tools and can be made by any logician—not only by narrow specialists of the pure or applied variety.

Let me begin with an example from "applied logic." As is well known, the United States and the Soviet Union almost share a common border. The Bering Strait cuts the two countries off where they very nearly touch. A little farther to the south, protruding from the southern end of Alaska and sprinkling themselves westward into the Pacific, are the Aleutian Islands. Off the Russian mainland (to be more exact, off the Kamchatka Peninsula) are also islands. What is not very well known, and has until recently been an official secret shared by Russia and the United States, is that lying amidst all these islands is one inhabited both by Russians and Americans.

Human beings being what they are, among the island's population are patriots and traitors. Now, it so happens that patriots always and only speak the truth to their compatriots, while traitors always and only speak the truth to those of the opposite nationality. To avoid confusion, both countries have agreed to allow no

visitors of any third nationality to the island, be they pro-Soviet, pro-American, or neutral.

Thus far I have merely recited a few geopolitical facts, interesting in themselves, but of no particular logical significance. Indeed, it might even appear to be a situation in which the special talents of the mathematical logician are of no particular relevance. Appearances are, however, often illusory. That mathematical logic can profitably be brought to bear on the subject becomes apparent by reflecting on the following:

**Gedankenexperiment 1.** Imagine yourself visiting this island. While wandering around you happen across a native. As the island is very far north, it can get quite cold and the native has buried himself under so many layers of clothing that you cannot judge by the costume whether he is an American or a Russian. A bit thoughtlessly you ask, "Are you an American or a Russian?" The native looks at you, spots your nationality immediately, replies, "I am a Russian," and leaves before you realize that you still have no idea if he is Russian or American. As you stand there and think about the situation for a while, it suddenly occurs to you that you do know whether the native was a patriot or a traitor. Which was he?

The most rudimentary ability to reason will solve the above puzzle. It is not the example of applied logic I spoke of. Indeed, although the island I referred to does exist and the Americans and Russians on the island have achieved the accommodation described, the puzzle itself is just a *Gedankenexperiment*: it does not describe reality and, therefore, is not *applied* logic. The application comes with the logical analysis of how Americans and Russians reason in solving this puzzle. This analysis yields the following:

**Amazing Fact.** Americans and Russians give different solutions to the puzzle of Gedankenexperiment 1.

The proof of this is quite simple and I omit it.

The interest in this fact is not so much in the details of the proof but in that it has a proof at all. One might compare it to the proof of the existence of Feigenbaum's number 4.6692016...; this important physical constant can be shown to exist and, indeed, calculated to any desired degree of accuracy—all by pure thought. In the case at hand, one might suspect on empirical grounds—e.g., the general disagreement between Russia and the United States on major international issues—that Russians and Americans reason differently. Indeed, the disparity of world views of the two nations makes such a difference not merely plausible, but even probable. And now, through pure thought, mathematical logic provides a proof of this difference.

Whenever one problem is solved, another generally arises. We know that Americans and Russians reason in different manners. Which of the two is more rational?

**Gedankenexperiment 2.** Imagine yourself on the island again. This time a native approaches you and says, "If I'm an American traitor or a Russian patriot, then you are an American." What should you make of this?

Given our first Gedankenexperiment, the following fact should no longer amaze the reader.

**Fact.** Americans will have no difficulty seeing that the native is telling the truth; a Russian will find the situation paradoxical.

We can see this quite easily. An American will reason as follows: "Whatever this fellow is, he has asserted a sentence of the form '$p$ implies $q$,' where $q$ is true. Therefore, he is telling the truth." A Russian, on the other hand, will reason along the following lines: "If this fellow is an American traitor or a Russian patriot, he will tell me the truth. But then his statement is of the form '$p$ implies $q$' with $p$ true and $q$ false, which cannot be. Therefore, this fellow is either an American patriot or a Russian traitor and is asserting an implication of the form '$p$ implies $q$' with $p$ false. But that means he is telling me the truth and must be an American traitor or a Russian patriot, which I have already ruled out. This makes no sense."

The discovery of such satisfying truths is, in pure science, an end in itself. (This is not so in applied science, where the goal is often to effect some change in the world. When one dips into applications, however much a purist one may be, one has a responsibility to apply one's results to the betterment of mankind. In the present case, one must analyze the implications that our knowledge of Russians' inferior mental abilities should have on such issues as negotiating treaties, funding scientific exchanges, and so forth. As this is hardly the place for me to elaborate on these matters, I refer the reader to my forthcoming monograph, entitled *Beyond Detente; A Puzzle Guide to Geo-Political Realities*, where such things are worked out in detail.)

Now I would like to shift gears, as it were, and consider a piece of pure logic. The point is, of course, that the same logical tools used in applied logic are useful in pure logic. In fact, the technique I applied above originated in highly abstruse, pure logic. To be specific, the technique comes from that branch of logic known as logical puzzle theory—or, more popularly, puzzle analysis—which was invented by Raymond Smullyan, whose monographs on the subject have been widely read and translated into several languages.

Smullyan's latest book, *Forever Undecided*, applies the techniques of puzzle analysis to Gedankenexperiments on natives of an imaginary island. Because he is able to create his own island, Smullyan is able to make the simplifying assumption that there are only two types of natives—knights, who always tell the truth, and knaves, who always lie.

**Gedankenexperiment 3.** Suppose you visit Smullyan's island and a native says to you, "You will never believe I'm a knight." Suppose, whether you realize it or not, you believe only true assertions. Is the native a knight or a knave? Will you ever believe the answer?

The answer is that (i) the native is a knight, (ii) you won't believe he is a knight, but (iii) you won't falsely believe he is a knave. Thus, there is an assertion which you will neither believe nor disbelieve (believing the native to be a knave amounts to believing him not to be a knight), i.e., you will remain forever undecided about this assertion.

The proof is fairly simple: if the native were a knave, he'd be lying and you would eventually believe he was a knight. Since you only believe true things, this cannot be the case. Thus, the native is a knight. Thus, what he says about your never believing him to be a knight is true. As for your never believing him to be a knave, this follows once again from the fact that you only believe true things.

A small change in the assumption of the Gedankenexperiment yields a different conclusion. If, for example, you believe that all your beliefs are true, then you

believe some false things. In fact, if you believe you are consistent, then you must believe yourself to be inconsistent!

A few words about interpretation: This is pure logic, not applied. It falls short of an application to epistemology because of the imaginary nature of the island and its oversimple picture of human nature. It does, however, apply to pure logic. This Gedankenexperiment can be modelled in formal systems containing some arithmetic to yield Gödel's First Incompleteness Theorem (any true formal theory containing enough arithmetic is incomplete); and the modified form yields the Second Incompleteness Theorem (any consistent formal theory containing enough arithmetic cannot prove its own consistency). Indeed, one may view this monograph not so much as a treatise on puzzle analysis as an exposition of Gödel's theorems and related results.

Inevitably, there arises the question of comparing Smullyan's exposition of Gödel's theorems to the other leading expositions. Here goes: The slim volume, *Gödel's Proof*, by E. Nagel and J. R. Newman can be read very quickly, but their wavering between presenting the details of coding and not presenting the details tends to mystify the proof rather than to de-mystify it. Moreover, the situation is not helped by their unconvincing philosophical remarks. Smullyan avoids both pitfalls by (i) omitting all the details of the arithmetization, and (ii) avoiding philosophy altogether.

If Smullyan beats Nagel and Newman, hands down, the contest with Douglas Hofstadter's *Gödel, Escher and Bach* and Rudolph von Bitter-Rucker's *Infinity and the Mind* is far from decided. These two books include the philosophy and the details, and much, much more, Hofstadter even including the kitchen sink. On the other hand, Hofstadter's book does run 777 pages, and both of these books offer rather more speculation than many professionals find comfortable.

In a nutshell, I would recommend Smullyan for high school students and Hofstadter and Rucker for more advanced students and nonspecialist professionals. I am, as already noted, less pleased with Nagel and Newman, and, in any event, their booklet is now out of date.

I shall finish with a couple of mild criticisms of Smullyan's book. On page 110, he says, "I, of course, am certainly all for popularization, providing the popularization is not inaccurate." Yet on p. 84 he says that he is not sure it is possible for a person to be peculiar in the sense that one can believe a proposition $p$ and yet believe that one doesn't believe $p$. This is exactly the sort of error that the insularization of purists and applicists causes to arise. Any logician with an eye toward applications would quickly cite Baire, Borel, and Lebesgue as examples: These men surely believed the axiom of choice—they used it readily enough; yet they believed they didn't believe it—they publicly declared it to be false.

Again, because of his snobbishly-purist attitude, Smullyan has overlooked important epistemological applications. Had he referred to a *real* island, like the one of Gedankenexperiments 1 and 2, he would have asked himself important (still open) questions like: Is there a proof of Gödel's Theorem that Americans understand but Russians find paradoxical? Is there a proof of Gödel's Theorem that Russians find paradoxical, but the proof that they find it paradoxical is itself paradoxical to the Americans?

# TELEGRAPHIC REVIEWS

*Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.*

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

| | | |
|---|---|---|
| T: Textbook | P: Professional Reading | 1–4: Semesters |
| C: Computer Software | L: Undergraduate Library | **: Special Emphasis |
| S: Supplementary Reading | 13: Grade Level | ??: Questionable |

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

**General, S*(13-18), L*.** *The Red Book: 100 Practice Problems for Undergraduate Mathematics Competition.* Kenneth S. Williams, Kenneth Hardy. Integer Pr, 1988, ix + 176 pp, $20 (P). [ISBN: 0-9692193-1-8] A companion to the authors' excellent first volume *The Green Book,* featuring instructive Putnam-like problems for undergraduates, compiled with hints and solutions. LCL

**General, T(13: 2).** *Mathematics for Business and Social Sciences, Fourth Edition: An Applied Approach.* Abe Mizrahi, Michael Sullivan. Wiley, 1988, xix + 875 pp, $47.38. [ISBN: 0-471-85291-0] Improvements include rewriting and reorganization of key chapters, several new examples and exercises. (*First Edition,* TR, December 1976; *Second Editon,* TR, August-September 1979; *Third Edition,* TR, January 1984.) LCL

**General, T.** *Mathematical Puzzling.* A. Gardiner. Oxford U Pr, 1987, 157 pp, $15.95 (P). [ISBN: 0-19-914258-0] An excellent collection of mathematical recreations (problems of discovery, exploration, and investigation) designed to highlight the true nature of mathematical thinking. Problems are divided into twenty-nine sections (arithmetic, geometry, counting, configurations, etc.), each followed by hints and comments. The material is especially appropriate for primary and secondary teachers wishing to counter the seemingly endless sequence of one-step problems that are so often set. LCL

**General, P*.** *Responses to the Challenge: Keys to Improved Instruction.* Ed: Bettye Anne Case. MAA Notes No. 11. MAA, 1989, vi + 266 pp, (P). [ISBN: 0-88385-061-3] A compilation of information on policies, practices, successes, failures, and goals connected with the use of teaching assistants and part-time instructors. Presents and analyzes data on who these teachers are, how they are used, and how they are being integrated into faculties. Includes helpful information, proven good advice, and useful material for both new and experienced college teachers. A project carried out by the joint MAA and AMS Committee on Teaching Assistants and Part-Time Instructors. CEC

**General.** *The Balomenos Lectures: Mathematics in Society and the Curriculum.* Ed: D.H. Van Osdol. U of New Hampshire (Dept. of Math., Durham, NH 03824), 1988, x + 53 pp, (P). Edited transcripts of three undergraduate colloquium lectures by Arthur Jaffe, Stephen Vardeman, and Henry Pollak followed by discussion by Philip J. Davis and Gian-Carlo Rota from a memorial lecture series held in Spring, 1987 at the University of New Hampshire in honor of the late Richard Balomenos. Stimulating ideas on educational implications of the changing state of mathematical science. LAS

**Elementary, T.** *Algebra for College Students.* James W. Hall. Ser. in Math. Prindle, Weber & Schmidt, 1988, xiv + 697 pp, $27. [ISBN: 0-87150-191-0] The topics of intermediate algebra (quadratic equations, systems of equations, exponential and logarithmic functions) for college students who have had one year of high school algebra. Aims to promote problem solving by including more word problems. Other priorities are exercises to improve estimation skill, and the integration of geometry whenever possible to promote visual reinforcement. Gobs of exercises, mastery tests, supplementary software, transparency masters. LCL

**Elementary, S.** *Ratio and Proportion in Mathematical Modelling.* B.M. Saler (30 Melva Crescent, Agincourt, Ontario, Canada M1V 1A3), 1985, 111 pp, $20 (P). Typed, xeroxed, plastic-bound workbook chockfull of examples to show how ratio and proportion are used in business, science, and industry. Seems overpriced. LCL

**Elementary, T(13: 1, 2).** *Contemporary Business Mathematics.* Ignacio Bello. DC Heath, 1988, xvii + 781 pp, $26. [ISBN: 0-669-12174-6] An assault on the concept of ratio and proportion: 700+ pages, including careful attention to objectives, motivations (color, format, pictures, diagrams, charts,

special vignettes "Minding Your Business" designed to make things relevant, attention to calculator and computer possibilities), reviews, and self-tests. Five parts: review of basics (arithmetic, equations, percent, and graphs, 150 pages), retail mathematics, accounting mathematics, mathematics of finance, and selected topics (statistics, the metric system, and binary arithmetic). LCL

**Mathematics Appreciation, T(13-15: 1), S, L**.** *Hidden Connections, Double Meanings.* David Wells. Cambridge U Pr, 1988, 164 pp, $14.95 (P). [ISBN: 0-521-31334-1] A collection of geometric problems that can be solved by "seeing" properly with the mind's eye—ranging from optical illusions to geometric proofs, from designs and patterns to analogies and metaphors. No algebra is used—only arithmetic and geometry. Many good problems, with hints (in one section) and solutions (in another). LAS

**Precalculus, T, S(13).** *Vectors.* B.M. Saler (30 Melva Crescent, Agincourt, Ontario, Cánada M1V 1A3), 1985, 103 pp, $20 (P). Typed, xeroxed, plastic-bound workbook features a leisurely introduction, with plenty of examples, to vector algebra (scalar and vector products) and applications, including equations of lines and planes. Note price! LCL

**Precalculus, S(13), L.** *Solving Problems in Algebra and Trigonometry.* V. Litvinenko, A. Mordkovich. MIR (US Distr: Imported Pub), 1987, 312 pp, $8.95. Over 2000 examples, problems, and exercises set forward here to develop a consummate skill in algebraic and trigonometric manipulation. An anachronism, but teachers may find it useful for students training for mathematical competitions. LCL

**Finite Mathematics, T(13).** *Finite Mathematics.* B.M. Saler (30 Melva Crescent, Agincourt, Ontario, Canada M1V 1A3), 1988, 304 pp, $50 (P). Three parts: 1) "Systems of linear equations and inequalities" covers matrices, linear programming, and the simplex method; 2) "Combinatorics" includes the usual counting techniques, combinations, and permutations; and 3) "Probability and statistics" treats the binomial, hypergeometric, normal and uniform distributions, and Markov chains. Individual sections available separately for $20 each. This typed, xeroxed, plastic-bound package is overpriced. LCL

**History, L**** *Introduction to Analysis of the Infinite, Book I.* Leonard Euler. Transl: John D. Blanton. Springer-Verlag, 1988, xv + 327 pp, $49.95. [ISBN: 0-387-96824-5] "Those things which are absolutely required for analysis" presented by Euler "so that the reader gradually and almost imperceptibly becomes acquainted with the idea of the infinite." 380 brief sections on infinite series, infinite products, and continued fractions. The first English translation of this classic work, which bridged what Euler believed to be a very considerable gap separating algebra from analysis. LAS

**History, S*, P, L*.** *I Want To Be A Mathematician: An Automathography in Three Parts.* Paul R. Halmos. MAA, 1985, xvi + 421 pp, $18 (P). [ISBN: 0-88385-445-7] Paperback reprint in MAA's new Spectrum Series of the 1985 Springer-Verlag edition (TR, August-September 1985). LAS

**History, P, L*.** *Srinivasa Ramanujan (1887-1920): A Tribute.* Ed: K.R. Nagarajan, T. Soundararajan. Macmillan India, 1988, xi + 96 pp, (P). A collection of historical and expository papers prepared on the occasion of the centenary of Ramanujan's birth. Opens with biographical sketches by K.S. Rao and G. Andrews; subsequent papers discuss Ramanujan's notebooks, the Rogers-Ramanujan identities, and his work on special functions. Cover features the bust of Ramanujan arranged by Richard Askey and executed by Paul Grunlund. LAS

**Logic, P.** *Stability in Model Theory.* Daniel Lascar. Transl: J.E. Wallington. Mono. & Surveys in Pure & Appl. Math., V. 36. Longman Scientific & Technical (US Distr: Wiley), 1987, 193 pp, $77.95. [ISBN: 0-582-99463-2] Concise presentation of stability theory ending with results on number of models for presentable, superstable theories. Assumes knowledge of first course in model theory. KS

**Logic, S, P, L.** *Matrix Logic.* August Stern. North-Holland (US Distr: Elsevier Science), 1988, viii + 215 pp, $68.50. [ISBN: 0-444-70432-9] An introduction to the matrix formulation of logic developed by the author, which not only allows one to derive the classical results of conventional logic but also permits their generalization. Boolean logic connectives are interpreted as matrix operators, and in this way matrix algebra logic extends the power of computation to the domain of logic valuations. Self-contained introduction to matrix algebra and symbolic logic, followed by synthesis and applications (especially to computer science). LCL

**Logic, T(18: 1, 2), P.** *Fundamentals of Stability Theory.* John T. Baldwin. Perspect. in Math. Logic. Springer-Verlag, 1988, xiii + 447 pp, $60. [ISBN: 0-387-15298-9] The aim of stability theory is to find invariants, if they exist, which determine the structure of a member of a class of algebras up to isomorphism. For example, dimension is an invariant for the class of vector spaces. Text presents axiomatic approach to notions of independence and generation, develops local dimension theory, and applies results to count number of models of a theory. Includes exercises and extensive bibliography. KS

**Logic, P, L.** *An Introduction to Independence for Analysts.* H.G. Dales, W.H. Woodin. Math. Soc. Lect. Note Ser., V. 115. Cambridge U Pr, 1987, xiii + 241 pp, $29.95 (P). [ISBN: 0-521-33996-0] The purpose of this book is to explain how the method of forcing yields independence and consistency results and to exemplify this by the study of an example from analysis. It is addressed to non-logicians, and in particular to analysts, who will find that foundational issues do arise naturally in analysis and that the technicalities of forcing are accessible. LCL

**Foundations, S(16-18), P.** *Logic, Methodology and Philosophy of Science VII.* Ed: Ruth Barcan Marcus, Georg J.W. Dorn, Paul Weingartner. Stud. in Logic & the Foundat. of Math., V. 114. North-Holland (US

Distr: Elsevier Science), 1986, xiii + 738 pp, $125. [ISBN: 0-444-87656-1] Proceedings of the Seventh International Congress, Salzburg, Austria, July 1983. The 41 invited papers cover a vast area including logic, mathematics, linguistics, natural sciences and humanities, e.g., Aspects of $\aleph_0$-Categoricity by G.L. Cherlin, and Conceptual Evolution and the Eye of the Octopus by D.L. Hull. MU

**Foundations, S, P, L.** *The World within the World.* John D. Barrow. Clarendon Pr, 1988, xiv + 398 pp. [ISBN: 0-19-851979-6] A collage of historical, philosophical, scientific, metaphorical, and personal speculations on the laws of nature, on the unspoken assumptions of scientific practice: that the universe is ordered, logical, mathematical, predictable, and in harmony with the working of our mind. Written for lay readers, this book focuses not on science but on meta-science: it addresses in countless ways the question of *why* nature's laws are as they seem to be. LAS

**Graph Theory, P, L.** *Selected Topics in Graph Theory 3.* Ed: Lowell W. Beineke, Robin J. Wilson. Academic Pr, 1988, x + 210 pp, $25. [ISBN: 0-12-086203-4] A collection of expository articles on various areas of graph theory of current interest. Authors include Ronald Graham, Fan Chung, Claude Berge, William Tutte, Lowell Beineke, and Robin Wilson. Topics include chromatic polynomials, matroids, hypergraphs, and nowhere-zero flow problems. LC

**Graph Theory, T(13: 1), L.** *Graph Theory: Euler's Rich Legacy.* Wayne Copes, *et al.* Contemp. Appl. Math. Janson, 1987, 72 pp, $7.95 (P). [ISBN: 0-939765-09-8] Readable monograph leads student through a gentle introduction of graph theory and its applications. Covers chromatic numbers, planarity, trees, and directed graphs. Plenty of examples, illustrations, and applications. Suitable for independent study. LC

**Combinatorics, P, L.** *Enumerative Combinatorics of Young Tableaux.* Shreeram S. Abhyankar. Pure & Appl. Math., V. 115. Marcel Dekker, 1988, xvii + 509 pp, $99.75. [ISBN: 0-8247-7816-2] Young tableaux were originally introduced in the context of invariant theory and later in the study of the irreducible representations of the symmetric group. During the last few years they have found application in many other areas of mathematics including algebraic geometry, statistics, and computer science. This self-contained account develops many enumerative formulas for tableaux and gives several applications. LCL

**Discrete Mathematics, P.** *Applications of Discrete Mathematics.* Ed: Richard D. Ringeisen, Fred S. Roberts. SIAM, 1988, x + 230 pp, $38.50. [ISBN: 0-89871-219-X] Proceedings of the Third SIAM Conference on Discrete Mathematics held on the campus of Clemson University. This volume consists of papers of many of the invited speakers and speakers at the minisymposia. Also includes a list of open problems from the problem session. CEC

**Discrete Mathematics, P.** *Lecture Notes in Computer Science-311: Coding Theory and Applications.* Ed: G. Cohen, P. Godlewski. Springer-Verlag, 1988, xiv + 196 pp, $21.80 (P). [ISBN: 0-387-19368-5] Twenty papers from the colloquium "Trois Journées sur le Codage," held in Cachan near Paris, France during November 24-26, 1986. A broad spectrum of recent work in information and coding theory. In English. CEC

**Number Theory, P.** *Applications of Fibonacci Numbers.* Ed: A.N. Philippou, A.F. Horadam, G.E. Bergum. Kluwer Academic, 1988, xx + 213 pp, $79. [ISBN: 90-277-2673-6] Proceedings of the Second International Conference on Fibonacci Numbers and Their Applications held August 1986 at San Jose State University. Nineteen research articles on the mathematics of recursive sequences. GG

**Number Theory, P, L*.** *Srinivasa Ramanujan: The Lost Notebook and Other Unpublished Papers.* Srinivasa Ramanujan. Springer-Verlag, 1988, xxv + 419 pp, $59.50. [ISBN: 0-387-18726-X] This volume includes reproductions of the "Lost Notebook" which consist of 90 pages of Ramanujan's work on $q$-series and other topics, work on properties of $\rho(n)$ and $\tau(n)$, 28 sheets from the "Loose Papers," and 117 pages of unpublished work related to various papers. Also includes reproductions of several letters between Ramanujan and some of his colleagues. A fascinating volume. CEC

**Number Theory, P.** *Ramanujan Revisited.* Ed: George E. Andrews, *et al.* Academic Pr, 1988, xx + 609 pp, $49.50. [ISBN: 0-12-058560-X] At the University of Illinois in June, 1987, 125 mathematicians gathered to commemorate the hundreth anniversary of Ramanujan's birth. This is the proceedings of that conference. Includes work related to Ramanujan's and also biographical and survey talks. CEC

**Number Theory, T(15-16: 1), S, L.** *A Course in Number Theory.* H.E. Rose. Clarendon Pr, 1988, xi + 354 pp, $32.50 (P); $59.95. [ISBN: 0-19-853261-X] A substantial introductory course with eight units on divisibility, congruence, algebraic numbers, continued fractions, quadratic forms, partitions, prime numbers, and diophantine equations. All but one of the units appear in two-chapter pairs including advanced material "best explored in the quiet of one's own study." Includes solutions to all problems in the back. LAS

**Linear Algebra, P.** *Linear Algebra in Signals, Systems, and Control.* Ed: Biswa Nath Datta, *et al.* SIAM, 1988, xiii + 667 pp, $58.50. [ISBN: 0-89871-223-8] Proceedings of the interdisciplinary SIAM conference of the same title, Boston, August 1986. 46 research papers organized into four categories: core linear algebra; numerical linear algebra; algorithms for signals, systems, and control; linear and nonlinear control and systems theory. RB

**Algebra, T(18), P.** *Ring Theory, Volume II.* Louis H. Rowen. Pure & Appl. Math., V. 128. Academic Pr, 1988, xiv + 462 pp, $84. [ISBN: 0-12-599842-2] An exposition of contemporary aspects of ring theory and algebra. The primary topics are homology

and cohomology, polynomial identities, central simple algebras, and representation theory. The writing is thorough; several exercises are provided; many interesting comments are included; and a large bibliography is attached. SG

**Algebra, P.** *Lecture Notes in Mathematics-1320: Semigroups: Theory and Applications.* Ed: H. Jürgensen, G. Lallement, H.J. Weinert. Springer-Verlag, 1988, x + 416 pp, $37.10 (P). [ISBN: 0-387-19347-2] A collection of twenty-eight papers covering such topics as consequences, varieties, languages, inverse semigroups, and semigroups of endomorphisms. SG

**Calculus, T*(13: 2, 3).** *Calculus with Analytic Geometry, Third Edition.* Howard Anton. Wiley, 1988, xxviii + 1377 pp, $47.50. [ISBN: 0-471-85045-4]. *Brief Edition.* xviii + 945 pp, $42. [ISBN: 0-471-62742-9] Changes in this edition include continuity being discussed before differentiability, differentiation has been reorganized, the chain rule has been rewritten, velocity and acceleration is presented more cohesively, logarithms have been rewritten, material on vector calculus has been rewritten, and a new chapter on second order differential equations has been added. Also includes more middle-level and hard problems. *Brief Edition* does not include chapters on vector-valued functions, partial derivatives, multiple integrals, topics in vector calculus, and three-dimensional space. A well-written, usable text. (*First Edition,* TR, April 1981; *Second Edition,* TR, November 1984; Extended Review, March 1986.) CEC

**Calculus, T(13: 2).** *Calculus with Trigonometry and Analytic Geometry.* John Saxon, Frank Wang. Saxon Publishers, 1988, x + 622 pp. [ISBN: 0-939798-34-4] The capstone of Saxon's innovative and controversial high school text series featuring problem sets that continually review all previous work, full use of informal differential notation, and heavy emphasis on specific devices to aid student calculation (e.g., a whole lesson on integrating even powers of sin and cos, and a two-page illustrated discussion of how to use fingers to hide parts of partial fractions to aid in finding the unknown coefficients). Content is entirely classical, aimed at helping students master techniques which, the authors believe, will slowly lead to understanding. No mention or use of calculators or computers, nor are there any "real" applications; problems are isomorphic to the least imaginative in every other calculus book, albeit sequenced in a dramatically different way. If your goal is to raise student scores on template tests, this is a way to do it. LAS

**Calculus, T(13).** *College Mathematics and Calculus with Applications to Management, Life and Social Sciences.* Karl J. Smith. Brooks/Cole, 1988, xvii + 729 pp, $32. [ISBN: 0-534-08910-0] Covers finite mathematics and calculus in one book. Modeling and applications are emphasized throughout. Finite mathematics portion includes matrices, linear programming, and probability. Calculus portion covers the standard topics with an emphasis on numerical and approximation technqiues. Techniques of integration are down-played. In the end, however, this text might be one which covers too much, too fast. MR

**Real Analysis, T(15-16), L.** *Introduction to Real Analysis.* John DePree, Charles Swartz. Wiley, 1988, xi + 355 pp. [ISBN: 0-471-85391-7] Recognizes the trend to merge the traditional advanced calculus and the first real variables course. Begins with real numbers and sequences, moves quickly to $R^n$ for half of the book, then on to complete metric spaces for concepts of compactness, connectivity, and the Stone-Weierstrass Theorem. The gauge integral follows quite naturally from the Riemann integral, but permits covering the Lebesgue theory. Attractive book. AWR

**Differential Equations, P.** *Lecture Notes in Mathematics-1338: Non-Oscillation Domains of Differential Equations with Two Parameters.* Angelo B. Mingarelli, S. Gotskalk Halvorsen. Springer-Verlag, 1988, xi + 109 pp, $13.10 (P). [ISBN: 0-387-50078-2] A differential equation is disconjugate (non-oscillatory) if every solution has at most one (finitely many) zeros. This is a study of the topological properties of the set of parameter values for which a certain two parameter family of linear second order differential equations is disconjugate or non-oscillatory. Vector equations and integral equations are also considered. Applications to Sturm-Liouville. SP

**Differential Equations, T(14: 1).** *A First Course in Differential Equations with Applications, Fourth Edition.* Dennis G. Zill. PWS-Kent, 1989, xii + 484 pp. [ISBN: 0-534-91568-X] The scope and emphasis of the original edition (*First Edition,* TR, August-September 1980) remain the same, but new features of this edition include expanded biographical footnotes, many new problems and examples, a new section on LRC circuits, and a revised treatment of complex eigenvalues for linear systems. A table of integrals has been included in the endpapers. (*Second Edition,* TR, March 1983; *Third Edition,* TR, February 1987.) CE

**Differential Equations, P.** *Matched Asymptotic Expansions: Ideas and Techniques.* P.A. Lagerstrom. Appl. Math. Sci., V. 76. Springer-Verlag, 1988, xii + 250 pp, $39.95. [ISBN: 0-387-96811-3] Discussion of selected topics and techniques in the use of matched asymptotic expansions for singular perturbation problems. Focus is on development of concrete examples, rather than rigor and completeness. Topics include classical model equations for fluid mechanics, layer type problems. RM

**Differential Equations, T(14: 1, 2).** *Differential Equations with Boundary-Value Problems, Second Edition.* Dennis G. Zill. PWS-Kent, 1989, xv + 671 pp. [ISBN: 0-534-91576-0] *Second Edition* of the expanded version of the *Fourth Edition* of the author's *A First Course in Differential Equations with Applications.* Some material has been rearranged,

some amplified, and some rewritten to improve clarity. Exercises and problems have been added. Student supplement available. Instructors may obtain complete solutions manual from the publisher. A middle-of-the-road book well worth a look. (*First Edition,* TR, February 1987.) JK

**Differential Equations, T*(16-17), L*.** *Difference Equations.* Ronald E. Mickens. Van Nostrand Reinhold, 1987, xii + 243 pp, $41.95. [ISBN: 0-442-26076-8] The purpose of this book is to present and explain mathematical methods for determining solutions to linear and nonlinear difference equations. The appealing writing style, the inclusion of many carefully worked examples, and the importance and relevance of the subject matter, makes this book a valuable addition to the undergraduate literature. No exercises. LCL

**Partial Differential Equations, P.** *Nonlinear Diffusion Equations and Their Equilibrium States I & II.* Ed: W.-M. Ni, L.A. Peletier, J. Serrin. Springer-Verlag, 1988. *I,* Math. Sci. Res. Inst. Publ., V. 12, xiii + 359 pp, $34.50 [ISBN: 0-387-96771-0]; *II,* Math. Sci. Res. Inst. Publ., V. 13, xiii + 365 pp, $35. [ISBN: 0-387-96772-9] Proceedings of a Microprogram held August 25-September 12, 1986 at the Mathematical Sciences Research Institute, University of Minnesota. Focus is on problems with equation $u_t = -u + f(u)$ and the equilibrium counterpart $-u + f(u) = 0$, where $f(u)$ is nonlinear. The fact that the proceedings fill two volumes attests to recent progress in a field once thought to be impossibly difficult. RBK

**Partial Differential Equations, P.** *Maximum Principles and Eigenvalue Problems in Partial Differential Equations.* Ed: P.W. Schaefer. Pitman Res. Notes in Math. Ser., V. 175. Longman Scientific & Technical (US Distr: Wiley), 1988, 230 pp, $54.95 (P). [ISBN: 0-470-21077-X] Proceedings of a conference held at the University of Tennessee, Knoxville, June 15-19, 1987, which focused on recent results and applications of maximum principles and the determination of bounds for the spectrum of elliptic partial differential equations and systems of equations. Five talks by M.H. Protter and eleven contributed lectures. RBK

**Partial Differential Equations, T(18), S, P.** *Variational Principles and Free-Boundary Problems.* Avner Friedman. Robert E Krieger, 1988, ix + 710 pp, $68.50. [ISBN: 0-89464-263-4] Corrected reprint of 1982 original edition published by Wiley (TR, March 1983). LAS

**Partial Differential Equations.** *Lecture Notes in Mathematics-1324: Partial Differential Equations.* Ed: F. Cardoso, *et al.* Springer-Verlag, 1988, viii + 433 pp, $37.10 (P). [ISBN: 0-387-50111-8] Partial differential equations research papers presented at the eighth Latin American School of Mathematics held in Rio de Janeiro, July 14-25, 1986. Emphasis on microlocal analysis, scattering theory and the applications of nonlinear analysis to elliptic equations and Hamiltonian systems. RBK

**Numerical Analysis, T(15-16: 1).** *Numerical Analysis.* Ian Jacques, Colin Judd. Chapman & Hall, 1987, vii + 326 pp, $34.50 (P); $67.50. [ISBN: 0-412-27960-6; 0-412-27950-9] An introductory textbook covering the solution of linear and nonlinear equations, the algebraic eigenvalue problem, data approximation, numerical integration, and the solution of ordinary differential equations. Solutions to *all* of the exercises are given in the back of the book. AO

**Numerical Analysis, P, L*.** *The Algebraic Eigenvalue Problem.* J.H. Wilkinson. Mono. on Numer. Analysis. Clarendon Pr, 1988, xviii + 662 pp, $39.95 (P). [ISBN: 0-19-853418-3] A paperback edition of this classic reference on numerical linear algebra. Text is identical to that in the 1978 hardcover printing. AO

**Numerical Analysis, P.** *Sequence Transformations.* Jean-Paul Delahaye. Ser. in Computat. Math., V. 11. Springer-Verlag, 1988, xxi + 252 pp, $79.50. [ISBN: 0-387-15283-0] Gives summary of sequence transformations and convergence acceleration. Covers accelerable and non-accelerable families of sequences, linear sequences, and automatic selection of sequence transformations. SP

**Numerical Analysis, S(15), L.** *Numerical Methods, with Applications in the Biomedical Sciences.* E.H. Twizell. Math. & Its Applic. Halsted Pr, 1988, 339 pp, $69.95. [ISBN: 0-470-21002-8] Written for undergraduate mathematics students and for those in biomedical sciences who need an introduction to numerical analysis. Chapters on linear and nonlinear systems, interpolation and approximation, differentiation and integration, and several chapters on differential equations. Few proofs of theorems, but many explanations, examples, and exercises. Calculus and, ideally, linear algebra are prerequisites. GG

**Numerical Analysis, P.** *Reliability in Computing: The Role of Interval Methods in Scientific Computing.* Ed: Ramon E. Moore. Perspectives in Comput., V. 19. Academic Pr, 1988, xv + 428 pp, $49.95. [ISBN: 0-12-505630-3] Papers from a series of lectures and software demonstrations given at the Ohio State University in 1987, on the theory and practice of interval arithmetic for reliable scientific computation. Papers on interval versions of programming languages (FORTRAN-SC, PASCAL-SC), architectures, floating point standards, algorithms, linear and nonlinear systems, optimization, and operator equations. RM

**Operator Theory, S(18), P.** *Operators and Representation Theory.* Palle E.T. Jorgensen. Math. Stud., V. 147. North-Holland (US Distr: Elsevier Science), 1988, viii + 337 pp, Dfl. 175.00 (P). [ISBN: 0-444-70321-7] Starting with Lie algebras, extensions, and projective representations, the author develops certain subjects from the theory of operator algebra and from representation theory. In some cases, this approach is new; also, the author shows that $C^*$-algebraic methods may be used with some success in nontraditional ways in the theory of

representations of infinite-dimensional Lie algebras. Throughout the book, the author stresses connections to mathematical physics. LW

**Analysis, P.** *Lecture Notes in Mathematics-1329: Orthogonal Polynomials and their Applications.* Ed: M. Alfaro, *et al.* Springer-Verlag, 1988, xv + 334 pp, $28.60 (P). [ISBN: 0-387-19489-4] Proceedings of the Second International Symposium on Orthogonal Polynomials and Their Applications at Sergovia, Spain, September 1986. LC

**Analysis, S(18), P.** *Geometric Inequalities.* Yu. D. Burago, V.A. Zalgaller. Transl: A.B. Sossinsky. Grund. der math. Wissenschaften, B. 285. Springer-Verlag, 1988, xiv + 331 pp, $97. [ISBN: 0-387-13615-0] A reference on inequalities for geometric characteristics of curves, surfaces, and other objects. A substantial part deals with isoperimeter inequalities and their generalizations. LCL

**Analysis, P.** *Marcel Riesz: Collected Papers.* Ed: Lars Gårding, Lars Hörmander. Springer-Verlag, 1988, vi + 897 pp, $65. [ISBN: 0-387-18115-6] Almost all the published papers, largely in French or German. JD-B

**Analysis, S(16-18), P, L.** *Descriptive Set Theory and the Structure of Sets of Uniqueness.* Alexander S. Kechris, Alain Louveau. London Math. Soc. Lect. Note Ser., V. 128. Cambridge U Pr, 1987, 367 pp, $34.50 (P). [ISBN: 0-521-35811-6] From lecture notes for a course taught at Caltech, UCLA; this is an unusual approach to the problem of when a trigonometric expansion is unique, a problem considered first by Riemann and Heine. MU

**Analysis, T(17-18: 1), P, L.** *Harmonic Analysis on Symmetric Spaces and Applications, II.* Audrey Terras. Springer-Verlag, 1988, xi + 385 pp, $45 (P). [ISBN: 0-387-96663-3] Nitty-gritty analysis on spaces of $n \times n$ matrices. Fewer applications than *Volume I.* Lots of exercises. An engaging style of exposition. BC

**Analysis, S, P*, L*.** *Inequalities, Second Edition.* G.H. Hardy, J.E. Littlewood, G. Pólya. Cambridge U Pr, 1988, xii + 324 pp, $19.95 (P). [ISBN: 0-521-35880-9] This classic, containing the statement and proof of all the standard inequalities of analysis, is available here as a paperback reprint, affordable for personal libraries. LCL

**Analysis, S(18), P.** *Poisson Algebras and Poisson Manifolds.* K.H. Bhaskara, K. Viswanath. Pitman Res. Notes in Math. Ser., V. 174. Longman Scientific & Technical (US Distr: Wiley), 1988, 128 pp, $47.95 (P). [ISBN: 0-582-01989-3] Assuming some familiarity with differentiable manifolds and Lie theory, the authors present a unified exposition on the current state of the theory of Poisson manifolds. With a unifying theme of Poisson algebras there are chapters on symmetric and alternating Schouten products, Poisson manifolds, and calculus on Poisson manifolds. Bibliography, index. JS

**Analysis, P.** *Mathematical Analysis and its Applications.* Ed: S.M. Mazhar, A. Hamoui, N.S. Faour. KFAS Proc. Ser., V. 3. Pergamon Pr, 1988, xi + 430 pp, $120. [ISBN: 0-08-031636-0] A selection of 38 invited talks and refereed research papers from a February 1985 international conference in Kuwait. LAS

**Algebraic Geometry, P.** *Lecture Notes in Mathematics-1335: Hyperrésolutions cubiques et descente cohomologique.* F. Guillén, *et al.* Springer-Verlag, 1988, xii + 192 pp, $20 (P). [ISBN: 0-387-50023-5]

**Algebraic Geometry, P.** *Lecture Notes in Mathematics-1311: Algebraic Geometry: Sundance 1986.* Ed: A. Holme, R. Speiser. Springer-Verlag, 1988, 320 pp, $30.30 (P). [ISBN: 0-387-19236-0] Proceedings of a conference at Sundance, Utah, July 1986. Most of the fourteen papers represent research actually begun or carried out at Sundance; the remainder consists of the usual writeups of lectures delivered at the conference. RB

**Algebraic Geometry, P.** *Lecture Notes in Mathematics-1337: Theory of Moduli.* Ed: E. Sernesi. Springer-Verlag, 1988, viii + 232 pp, $20 (P). [ISBN: 0-387-50080-4] Contains three long papers: F. Catanese, "Moduli of Algebraic Surfaces;" R. Donagi, "The Schottky Problem;" and J.L. Harer, "The Cohomology of the Moduli Space of Curves." SG

**Algebraic Geometry, P.** *Vector Bundles on Algebraic Varieties.* M.F. Atiyah, *et al.* Oxford U Pr, 1987, vi + 555 pp, $35 (P). A collection of nineteen papers presented at a conference in Bombay held in 1984. SG

**Algebraic Geometry, P.** *Equimultiplicity and Blowing Up: An Algebraic Study.* M. Herrmann, S. Ikeda, U. Orbanz. Springer-Verlag, 1988, xvii + 629 pp, $99.50. [ISBN: 0-387-15289-X] A research monograph in commutative algebra treating the behavior of Hilbert functions and Cohen-Macaulay rings during resolution (blowing up) of algebraic or complex analytic singularities. Includes an extensive appendix (150 pp) by B. Moonen on geometric equimultiplicity. LAS

**Algebraic Geometry, P.** *Lecture Notes in Mathematics-1310: Local Moduli and Singularities.* Olav Arnfinn Laudal, Gerhard Pfister. Springer-Verlag, 1988, v + 117 pp, $13.90 (P). [ISBN: 0-387-19235-2] The outgrowth of a collaboration between the authors over the last five years. The local moduli problem attempts to describe the set of isomorphism classes of algebraic objects which occur as the result of arbitrary small deformations of a given algebraic object such as a projective $k$-scheme. CEC

**Algebraic Geometry, P.** *Lecture Notes in Mathematics-1327: Determinantal Rings.* Winfried Bruns, Udo Vetter. Springer-Verlag, 1988, vii + 236 pp, $20 (P). [ISBN: 0-387-19468-1] A coherent treatment of determinantal rings building on the basics of commutative algebra as found, for example, in Part I of Matsumura's book. The main approach is via the theory with straightening law, but other methods have not been neglected. LCL

**Differential Geometry, P.** *Lecture Notes in Mathematics-1339: Geometry and Analysis on Manifolds.* Ed: T. Sunada. Springer-Verlag, 1988, ix + 277 pp, $24.30 (P). [ISBN: 0-387-50113-4] Proceedings of a symposium (Katata, Japan) and subsequent conference (Kyoto), August 1987. Seventeen papers on various aspects of geometric analysis, including spectral analysis of the Laplacian on compact and noncompact Riemannian manifolds, harmonic analysis on manifolds, complex analysis and isospectral problems. RB

**Geometry, P, L**.** *A History of Non-Euclidean Geometry: Evolution of the Concept of a Geometric Space.* B.A. Rosenfeld. Transl: Abe Shenitzer. Stud. in History of Math. & Physic. Sci., V. 12. Springer-Verlag, 1988, xi + 471 pp, $89. [ISBN: 0-387-96458-4] A readable, corrected, and supplemented translation of the original 1976 Russian edition. Discusses the mathematical and philosophical factors underlying the discovery of non-Euclidean geometry. Updated reference list of over 650 items includes books and papers that have appeared since the original date of publication. Additions to text include links between ancient and later Arabic mathematics, new publications on the history of the theory of parallel lines, a section on Apollonius' geometric transformations, a section on three-dimensional geometric algebra, plus other matters. For every college and university library. JK

**Algebraic Topology, P.** *Lecture Notes in Mathematics-1318: Algebraic Topology—Rational Homotopy.* Ed: Y. Felix. Springer-Verlag, 1988, viii + 245 pp, $25.80 (P). [ISBN: 0-387-19340-5] Proceedings of a meeting at Louvain-la-Neuve, Belgium, May 1986, organized in order to discuss and promote new developments in algebra and algebraic topology related to rational homotopy. Three surveys (Hilbert and Poincaré series, "tame" homotopy theory, Moore conjectures), thirteen research papers (five in French). RB

**Algebraic Topology, P.** *Lecture Notes in Mathematics-1326: Elliptic Curves and Modular Forms in Algebraic Topology.* Ed: P.S. Landweber. Springer-Verlag, 1988, 224 pp, $20 (P). [ISBN: 0-387-19490-8] Twelve papers, with both topologists and number theorists as authors, from a September 1986 conference at the Institute for Advanced Study. Elliptic genera and elliptic cohomology form central theme. Connection with algebraic geometry is through formal groups. GG

**Algebraic Topology, T**(18: 2), L.** *An Introduction to Algebraic Topology.* Joseph J. Rotman. Grad. Texts in Math., V. 119. Springer-Verlag, 1988, xiii + 433 pp, $49.80. [ISBN: 0-387-96678-1] Delightful writing that catches the interplay of geometry and algebra. There is an unusually large number of instructive exercises that, together with the text, cover the range from historical foundations to the categorical tools of homological algebra. Balances the study of homotopy and (co)homology. JAS

**Topology, T(16-18: 1, 2).** *Topologie II: Uniforme Räume mit zwei Zeichnungen von Volker Kühn.* Horst Herrlich. Berliner Studienreihe zur Math., V. 3. Heldermann Verlag, 1988, xi + 265 pp, $29 (P). [ISBN: 3-88538-103-6] The third in a series of university course lecture notes on topology, preceded by *Einführung in die Topologie: Metrische Räume* and *Topologische Räume.* This volume develops basic topological concepts in the framework of nearness spaces. Exercises, bibliography, and index make this more appropriate as a text than most "lecture notes." JAS

**Operations Research, P.** *Queueing Theory and its Applications: Liber Amicorum for J.W. Cohen.* Ed: O.J. Boxma, R. Syski. CWI Mono., No. 7. North-Holland (US Distr: Elsevier Science), 1988, xxii + 446 pp, $97.25. [ISBN:0-444-70497-3] Dedicated by friends and students to mark the retirement of J.W. Cohen, early contributor to the development of queueing theory, these papers are grouped in five areas: I: Broad Survey; II: Single Server; III: Analytic Methods; IV: Applications to Communication and Computing; V: Topics in Probability and Statistics. AWR

**Operations Research, T(17-18: 2), P, L.** *Integer and Combinatorial Optimization.* George L. Nemhauser, Laurence A. Wolsey. Wiley, 1988, xiv + 763 pp, $79.95. [ISBN: 0-471-82819-X] An expensive but extremely well-written graduate text covering integer programming. Substantial care is taken in the presentation of theoretical aspects of integer programming. Good exercises at the end of most chapters; no solutions. Very easy to read and digest. SM

**Optimization, T(17: 1, 2), L.** *Linear Complementarity, Linear and Nonlinear Programming.* K.G. Murty. Sigma Ser. in Appl. Math., V. 3. Heldermann Verlag, 1988, xlviii + 629 pp, $90 (P). [ISBN: 3-88538-403-5] A sequel to the author's book *Linear and Combinatorial Programming.* Provides an introduction to geometric, mathematical, and computational aspects of the linear complementarity problem and the quadratic programming problem. The final chapter surveys some recent results in linear programming (e.g., Karmarkar's algorithm). AO

**Optimization, T(15-17: 1), L.** *Introduction to Optimization.* E.M.L. Beale. Disc. Math. & Optimiz. Wiley, 1988, ix + 121 pp, $31.95. [ISBN: 0-471-91760-5] An introduction to several of the topics covered by the term "optimization:" unconstrained optimization, linear programming, nonlinear programming, integer programming, and dynamic programming. No exercises. AO

**Dynamical Systems, P.** *Lecture Notes in Mathematics-1309: Global Bifurcation of Periodic Solutions with Symmetry.* Bernold Fiedler. Springer-Verlag, 1988, viii + 144 pp, $17.30 (P). [ISBN: 0-387-19234-4] Treats the problem of finding periodic orbits in dynamical systems with symmetries by developing a generic global equivariant Hopf index. A global Hopf bifurcation is implied when this index is non-zero. Applications to coupled oscillators and

reaction diffusion systems. SP

**Dynamical Systems, P.** *Lecture Notes in Mathematics-1331: Dynamical Systems, Valparaiso 1986.* Ed: R. Bamón, R. Labarca, J. Palis Jr. Springer-Verlag, 1988, vi + 250 pp, $24.30 (P). [ISBN: 0-387-50016-2] Conference proceedings of the Second International School of Dynamical Systems held at the Universidad Tecnica Federico Santa Maria-Valparaiso, Chile. A collection of 15 research articles dealing with bifurcation and singularities of vector fields, Anosov flows, Hausdorff dimension, limit capacity, and foliations. CE

**Dynamical Systems, P*.** *Global Bifurcations and Chaos: Analytical Methods.* Stephen Wiggins. Appl. Math. Sci., V. 73. Springer-Verlag, 1988, xiv + 494 pp, $49.95. [ISBN: 0-387-96775-3] A rigorous treatment of chaotic phenomena in nonlinear dynamical systems written with the applied scientist in mind. Includes background material, definition of chaos, and lots of examples. Treats the Smale Horseshoe, symbolic dynamics, general criteria for chaos in hyperbolic and nonhyperbolic systems, homoclinic orbits, and Melnikov-like methods of detecting chaos. SP

**Dynamical Systems, P, L.** *Singularities and Groups in Bifurcation Theory, Volume II.* Martin Golubitsky, Ian Stewart, David G. Schaeffer. Appl. Math. Sci., V. 69. Springer-Verlag, 1988, xvi + 533 pp, $69.50. [ISBN: 0-387-96652-8] Bifurcation theory studies how the solutions to equations (algebraic, ordinary differential equations, partial differential equations) change their structure (topological, geometric) as parameters in the equations are varied. In this second volume of a comprehensive work, the authors focus on how the use of group theoretic techniques aids in understanding bifurcation of systems with symmetry, i.e., systems whose solution sets are invariant under an invertible linear transformation. Included are case studies of Benard convection, deformation of elastic material, and Taylor-Couette flow. CE

**Dynamical Systems, P.** *Lecture Notes in Mathematics-1294: Substitution Dynamical Systems—Spectral Analysis.* Martine Queffélec. Springer-Verlag, 1987, xiii + 240 pp, $25.80 (P). [ISBN: 0-387-18692-1] A complete and unified description of substitution dynamical systems arising from substitutions of constant length, linking several domains: combinatorics, ergodic theory, and harmonic analysis of measures. LCL

**Dynamical Systems, T(18: 2), P.** *Asymptotic Behavior of Dissipative Systems.* Jack K. Hale. Math. Surv. & Mono., No. 25. AMS, 1988, ix + 198 pp, $54. [ISBN: 0-8218-1527-X] A dynamical system is dissipative if there exists a bounded set (an attractor) into which every orbit eventually enters and remains. Here we have a presentation of the dissipative systems in the infinite dimensional case. A general class of systems, called asymptotically smooth, are given for which a global attractor $A$ exists. Large number of examples and applications. MR

**Dynamical Systems, P.** *Chaotic Motions in Nonlinear Dynamical Systems.* W. Szemplińska-Stupnicka, G. Iooss, F.C. Moon. Springer-Verlag, 1988, 193 pp, $33.10 (P). [ISBN: 0-387-82062-0] Contains three lectures of interest to applied scientists interested in chaotic phenomena. Presents viewpoints of both the mathematician and the experimenter in nonlinear dynamics. Discusses Poincaré maps, fractal dimension, Lyaponov exponents, the relation between chaos and classical perturbation techniques, and methods of calculating bifurcations. SP

**Control Theory, P.** *Control and Dynamical Systems: Advances in Theory and Applications, V. 27: System Identification and Adaptive Control, Part 3 of 3.* Ed: C.T. Leondes. Academic Pr, 1988, x + 377 pp, $65. [ISBN: 0-12-012727-X] Part of a trilogy intended as a comprehensive reference for practitioners in the field. Besides articles describing new general techniques (linear programming approach to constrained multivariable process control, techniques using finite element approximation), there are articles as specific as optimal control for air-conditioning systems in large buildings. AWR

**Control Theory, P.** *Lecture Notes in Control and Information Sciences-112: State-Space Models of Lumped and Distributed Systems.* V. Kecman. Springer-Verlag, 1988, ix + 280 pp, $30.90 (P). [ISBN: 0-387-50082-0] Example-oriented treatment of dynamic processes, such as mass storage (inflow and outflow), fluid flow, and heat transfer. A "lumped" process occurs simultaneously throughout a system; a "distributed" processes varies spatially. From the Preface: "The reader will find basic and indispensable knowledge here, and master the techniques and secrets for devising a dynamic mathematical model making use of the fundamental laws of nature." BC

**Control Theory, P.** *Lecture Notes in Mathematics-1313: Optimal Periodic Control.* Fritz Colonius. Springer-Verlag, 1988, vi + 177 pp, $17.30 (P). [ISBN: 0-387-19249-2] Terse notes concerned with optimal periodic control for ordinary and functional differential equations of retarded type. AWR

**Control Theory, P.** *Lecture Notes in Control and Information Sciences-105: Modelling and Adaptive Control.* Ed: Ch. I. Byrnes, A. Kurzhanski. Springer-Verlag, 1988, v + 384 pp, $48 (P). [ISBN: 0-387-19019-8] Proceedings of a 1986 workshop in Sopron, Hungary on dynamical system modeling and control in the face of uncertainty. RM

**Control Theory, T(18: 3), P.** *Linear Stochastic Systems.* Peter E. Caines. Prob. & Math. Stat. Wiley, 1988, xiv + 874 pp, $59.95. [ISBN: 0-471-08101-9] Mathematically-rigorous treatment of estimation and control theory for linear stochastic systems. Stochastic processes and systems, theory of prediction, system identification, parameter estimation, and control, and stochastic adaptive control theory. Ph.D. thesis, tested in graduate courses. TH

**Control Theory, P.** *Perturbation Methods in Op-*

*timal Control.* Alain Bensoussan. Modern Appl. Math. Wiley, 1988, xiv + 573 pp, $59.95. [ISBN: 0-471-91994-2] Authors distinguish their work from the very large literature on perturbation that focuses on differential equations rather than control theory, and from the work of J.L. Lions [*Problèmes aux Limites non Homogènes et Applications, V. 1, 2, 3,* Dunod, 1968, 1969] about control problems with perturbations for systems described by partial differential equations. This book collects from the literature control problems described by ordinary differential equations. AWR

**Probability, P.** *Lecture Notes in Mathematics-1321: Séminaire de Probabilités XXII.* Ed: J. Azéma, P.A. Meyer, M. Yor. Springer-Verlag, 1988, iv + 600 pp, $55.90 (P). [ISBN: 0-387-19351-0] Forty research articles plus errata from *Séminaire de Probabilités XX* (LNM 1204). GG

**Probability, T(15-16).** *Introduction to Stochastic Models.* Roe Goodman. Benjamin/Cummings, 1988, xii + 368 pp, $39.95. [ISBN: 0-8053-6011-5] An introduction to probability and stochastic processes aimed at juniors or seniors, but really presuming only one year of calculus and introductory linear algebra. Focus is on random variables and expectation rather than combinatorics, and covers many classic probability problems. Includes Markov chains, the Poisson process, and queueing. Attractive. AWR

**Probability, P.** *Unimodality, Convexity, and Applications.* Sudhakar Dharmadhikari, Kumar Joag-dev. Prob. & Math. Stat. Academic Pr, 1988, xii + 278 pp, $64.50. [ISBN: 0-12-214690-5] Theoretical monograph developing the concept of unimodality of a distribution (unique maximum value) through the convexity of the distribution function. Main applications are to statistical inference and reliability theory. RSK

**Stochastic Processes, P.** *Lecture Notes in Mathematics-1316: Stochastic Analysis and Related Topics.* Ed: H. Korezlioglu, A.S. Ustunel. Springer-Verlag, 1988, 371 pp, $34.80 (P). [ISBN: 0-387-19315-4] Twelve papers (10 English, 2 French) from a 1986 workshop in Silivri, Turkey. Three invited lectures on Brownian motion, diffusion, and infinite dimensional calculus; stochastic calculus of variations; and noncausal stochastic integrals and calculus. TH

**Stochastic Processes, T(18: 1), P, L.** *Numerical Techniques for Stochastic Optimization.* Ed: Yu. Ermoliev, R. J-B Wets. Ser. in Computat. Math., V. 10. Springer-Verlag, 1988, xiii + 571 pp, $74. [ISBN: 0-387-18677-8] A collection of articles covering numerical solution of stochastic programs. An introductory chapter is followed by seven chapters on numerical techniques of solution, twelve chapters on the computer implementation of these techniques, and nine chapters giving case studies and test problems. The typeface of this book appears to have been produced on a laser printer with poor resolution, making this book very tiring on the eyes. SM

**Elementary Statistics, T(13: 1).** *Exploring Statistics.* Damaraju Raghavarao. Stat.: Textbooks & Mono., V. 92. Marcel Dekker, 1988, ix + 276 pp, $39.75. [ISBN: 0-8247-7952-5] For self-study or non-mathematical introductory course covering data collection and summaries, simple inference, multivariate relationships, ANOVA, and Chi-square. Formulas (e.g., mean and standard deviation) are concentrated in chapter appendices; chapters concentrate on ideas and examples and avoid calculations. Some examples use Minitab. Latin square example is sexist. TH

**Computational Statistics, P, L.** *The New S Language: A Programming Environment for Data Analysis and Graphics.* Richard A. Becker, John M. Chambers, Allan R. Wilks. Wadsworth, 1988, xvii + 702 pp, $29.95 (P); $39.95. [ISBN: 0-534-09192-X] S is a language and an interactive programming environment for data analysis and graphics. The first half of this book is an introduction to the capabilities of S. The second half is a reference manual that contains detailed documentation of the S functions and datasets. An appendix describes the relationship between this version of S and an earlier version described in *S: An Interactive Environment for Data Analysis and Graphics* by R.A. Becker and J.M. Chambers, Wadsworth, 1984 (TR, February 1985). AO

**Statistics, P*.** *Goodness-of-Fit Statistics for Discrete Multivariate Data.* Timothy R.C. Read, Noel A.C. Cressie. Ser. in Stat. Springer-Verlag, 1988, xi + 211 pp, $44. [ISBN: 0-387-96682-X] Thorough analysis and comparison of goodness-of-fit statistics. Treatment is unified by using the authors' power-divergence family of statistics, which includes Pearson's $X^2$ and the log likelihood ratio statistic $G^2$ as special cases. Also provides a detailed historical account of these two statistics, and a chapter on future directions for research. Good bibliography. RSK

**Statistics, S*(14-18), P, L*.** *Encyclopedia of Statistical Sciences, Volume 9: Strata Chart to Zyskind-Martin Models, Cumulative Index, Volumes 1-9.* Ed: Samuel Kotz, Norman L. Johnson. Wiley, 1988, xii + 762 pp, $125; $775 (Set). [ISBN: 0-471-85474-3] At long last, the final volume in what turned out to be a decade-long project! (See TR, October 1982, of Series and *Volume 1.*) An important addition to the literature for the non-specialist. A supplementary volume is projected for the near future. Note special price for whole set. RSK

**Statistics, P.** *Foundations of Estimation Theory.* Lubomír Kubáček. Fund. Stud. in Engin., V. 9. Elsevier Science, 1988, vi + 328 pp, $117. [ISBN: 0-444-98941-2] Updated translation of a Czechoslovakian monograph. Theoretical presentation assuming a background of matrix theory, functional analysis, and advanced probability theory. Note price! RSK

**Statistics, T(17: 1, 2), P*.** *Applied Regression Analysis: A Research Tool.* John O. Rawlings. Stat./Prob. Ser. Wadsworth, 1988, xvii + 553 pp, $49.95. [ISBN: 0-534-09246-2] Non-theoretical treatment assuming a year course in statistical meth-

ods. First part reviews least squares regression using matrix notation, with a detailed case study demonstrating the techniques. Middle part contains chapters on the geometric interpretation of least squares, selection procedures, class variables in regression, problem areas, diagnostics, and transformations. Final part consists of chapters on collinearity, response surface modeling, and analysis of unbalanced data, with detailed case studies illustrating each. RSK

**Statistics, T(17-18: 1, 2), S, P.** *Linear Structures.* Jan R. Magnus. Stat. Mono., V. 42. Oxford U Pr, 1988, xii + 205 pp, $49.95. [ISBN: 0-85264-299-7] A linear structure is a class of matrices with some constraint: symmetric, triangular, diagonal, or skew-symmetric. A special case of such a structure is the commutation matrix which transforms the vector vec $A$ formed by stacking columns of $A$ one underneath the other into vec $A'$; also, the properties of linear structures can be used to obtain Jacobian matrices of transformations involving such structures. Developed from a course for graduate students in econometrics and statistics, the book presupposes "a modest general background in undergraduate mathematics and matrix algebra." Exercises follow each section. LW

**Computer Literacy, S, P, L.** *IBM: The Making of the Common View.* Michael Killen. Harcourt Brace Jovanovich, 1988, xx + 284 pp, $17.95. [ISBN: 0-15-143480-8] An overly-dramatized account of IBM's effort from 1983 to 1987 to bring consistency to their computer systems through System Application Architecture (SAA). Heavy on management analysis, light on computer details or philosophy. LAS

**Programming, S(14-16).** *Common Lisp Drill.* Taiichi Yuasa. Transl: Richard Weyhrauch, Yasuko Kitajima. Academic Pr, 1988, x + 232 pp, $19.95 (P). [ISBN: 0-12-774861-X] Extensive, well-designed set of exercises for common Lisp, based on the text *Introduction to Common Lisp* by the author and Masami Hagiya (TR, June-July 1988). Nicely coordinated with the earlier book, though one wonders why they were not included in it. RM

**Languages, S(14-17), P, L.** *Prolog by Example: How to Learn, Teach and Use It.* Helder Coelho, José C. Cotta. Symbolic Computat. Springer-Verlag, 1988, x + 382 pp, $35. [ISBN: 0-387-18313-2] The descendant of the authors' widely used "How to solve it with Prolog" (1979), this guide book for logic programming software development presents 175 problems with annotated solutions in the Prolog language. The learn-by-example pedagogical material graduates in difficulty from novice to professional, and is supplemented by brief chapters for teachers and students, and lists of Prolog implementations. RB

**Languages, S(14-16), P, L.** *C: An Advanced Introduction (ANSI C Edition).* Narain Gehani. Principles of Comput. Sci. Computer Science Pr, 1988, xv + 265 pp, $23.95 (P). [ISBN: 0-7167-8196-4] A version of the author's popular book on programming in C language, revised to agree with the nearly-approved ANSI standard language definition, with overviews of the superset languages C++ and Concurrent C; description of ANSI standard library; differences between ANSI and the prior *de facto* standard "K&R;" and an annotated bibliography. Presumes experience with another high-level language; focuses on advanced features. RB

**Languages, P.** *Lecture Notes in Computer Science-306: Foundations of Logic and Functional Programming.* Ed: M. Boscarol, L. Carlucci Aiello, G. Levi. Springer-Verlag, 1988, v + 218 pp, $21.80 (P). [ISBN: 0-387-19129-1] Ten papers from the proceedings of a workshop on logic and functional programming held in Trento, Italy, December 15-19, 1986. LCL

**Algorithms, P.** *Lecture Notes in Computer Science-318: SWAT 88.* Ed: R. Karlsson, A. Lingas. Springer-Verlag, 1988, vi + 262 pp, $23.10 (P). [ISBN: 0-387-19487-8] Proceedings of the 1st Scandanavian Workshop on Algorithm Theory held July 5-8, 1988 in Halmstad, Sweden. Contains the texts of four invited and twenty-four contributed papers describing research in various areas of algorithm theory: data structures, computational geometry, and computational complexity. AO

**Algorithms, T*(16-17: 1), P.** *Efficient Parallel Algorithms.* Alan Gibbons, Wojciech Rytter. Cambridge U Pr, 1988, viii + 259 pp, $59.50. [ISBN: 0-521-34585-5] Design of algorithms for parallel computation in the context of a high-level programming language rather than that of specific machine architectures. Studies problems with an inherently parallel nature concerning graphs, expression evaluation, recognition and parsing of context-free languages, sorting, strong matching. Mentions problems not apparently parallel in nature. Targeted at non-specialists. DFA

**Algorithms, T(14: 1).** *Fundamentals of Data Structures in Turbo Pascal: For the IBM PC.* Ellis Horowitz, Sartaj Sahni. Comput. Software Eng. Ser. Computer Science Pr, 1989, x + 478 pp, $39.95. [ISBN: 0-7167-8152-2] Adapts authors' popular *Fundamentals of Data Structures* for users of the IBM PC (or equivalent) and Turbo Pascal (Versions 3 and 4). Assumes a prior programming course. Introduces the chosen computer and language, then presents all the usual topics: arrays, stacks, queues, lists, trees, graphs, sorting methods, sets, game-playing, hashing. Emphasis on design of good structures and time/space analysis of the resulting algorithms. DFA

**Computer Systems, P.** *Creating User Interfaces by Demonstration.* Brad A. Myers. Perspectives in Comput., V. 22. Academic Pr, 1988, xxiii + 276 pp, $29.95. [ISBN: 0-12-512305-1] Describes the design and implementation of Peridot, a tool for creating complex, graphical, highly-interactive user interfaces by demonstration rather than programming. Integrates and incorporates the results of recent research in visual programming, programming by example, constraints, and plausible inference. Based

on the author's Ph.D. research at the University of Toronto. AO

**Computer Systems, P.** *Lecture Notes in Computer Science-309: Experiences with Distributed Systems.* Ed: J. Nehmer. Springer-Verlag, 1988, vi + 292 pp, $24.50 (P). [ISBN: 0-387-19333-2] Proceedings of an international workshop at Kaiserlautern, FRG, September 1987 in which (good and bad) experiences with existing concepts in distributed system design were exchanged. Seven reports on specific projects (Amoeba, CONIC, DAC, HCS, INCAS, POOL, PEACE) and five papers on experiences with topics accumulated through several projects (fault tolerance, two on design principles, two on distributed databases). RB

**Computer Systems, S(15-17), P.** *The Architectural Logic of Database Systems.* E.J. Yannakoudakis. Springer-Verlag, 1988, xiv + 318 pp, $39 (P). [ISBN: 0-387-19513-0] A general, example-oriented introduction to database principles and presentation of current database management software at various levels, without reference to vendor-related software. Intended for students, programmers, system analysts, database administrators and data processing managers. Presumes elementary knowledge of programming. Standard DBMS languages discussed; sizable dictionary of terms. RB

**Computer Systems, S, P, L*.** *Computer Algebra: Systems and Algorithms for Algebraic Computation.* J.H. Davenport, Y. Siret, E. Tournier. Transl: A. Davenport, J.H. Davenport. Academic Pr, 1988, xix + 267 pp, $14. [ISBN: 0-12-204230-1] The first systematic exposition (translated from the French original) of algorithms and data representation used in computer algebra systems. Covers polynomial simplification, modular and $p$-adic methods, formal integration, and ordinary differential equations. An appendix contains a complete description of REDUCE, one of the earliest computer algebra systems. LAS

**Computer Systems, P.** *Lecture Notes in Computer Science-312: Distributed Algorithms.* Ed: J. van Leeuwen. Springer-Verlag, 1988, vii + 430 pp, $33.30 (P). [ISBN: 0-387-19366-9] Proceedings of the second international workshop held in Amsterdam, July 8-10, 1987. JAS

**Theory of Computation, T(16-17: 1), S, L.** *Parsing Theory, Volume I: Languages and Parsing.* Seppo Sippu, Eljas Soisalon-Soininen. EATCS Mono. on Theoret. Comput. Sci., V. 15. Springer-Verlag, 1988, viii + 228 pp, $59.50. [ISBN: 0-387-13720-3] A textbook and up-to-date reference work on the theory of deterministic parsing of context-free grammars. The presentation treats the material in a mathematical spirit, deriving construction algorithms from general graph-theoretic methods, paying special attention to complexity issues. This volume introduces the basic concepts of languages and parsing, including mathematical and computer scientific background. Presumes elementary formal language theory, complexity, data structures, analysis of algorithms. Over 200 exercises. RB

**Theory of Computation, P.** *Computing in Horn Clause Theories.* Peter Padawitz. EATCS Mono. on Theoret. Comput. Sci., V. 16. Springer-Verlag, 1988, xi + 322 pp, $45. [ISBN: 0-387-19427-4] This monograph presents Horn logic with equality as a unifying common background behind semantical concepts, recursion, logic programming, term rewriting, data type specification and automated theorem proving. This approach provides for "natural" problem specification, multiple semantical views, and deductive methods usable for rapid prototyping. Self-contained; requires general ability to deal with formalisms, inductive proof. RB

**Theory of Computation, P.** *Lecture Notes in Computer Science-316: Automata Networks.* Ed: C. Choffrut. Springer-Verlag, 1988, vii + 125 pp, $15.40 (P). [ISBN: 0-387-19444-4] Proceedings of the annual University of Paris Spring School conference (May 1988) in which the "state-of-the-art" in a specific topic is surveyed. Nine papers: cellular automata, relationship with group theory, one-dimensional case, properties of finite automata networks (three papers), the "firing squad" problem, an application to artificial intelligence, a probabilistic approach. RB

**Theory of Computation, P.** *Lecture Notes in Computer Science-323: Attribute Grammars: Definitions, Systems and Bibliography.* Pierre Deransart, Martin Jourdan, Bernard Lorho. Springer-Verlag, 1988, ix + 232 pp, $20.60 (P). [ISBN: 0-387-50056-1] A survey of the present state-of-the-art in attribute grammars, in three parts. *Definitions* reviews results, promising research directions in terms of a unified vocabulary. *Systems* provides descriptions of over 30 software systems which implement attribute evaluation. Thematically indexed *bibliography.* For researchers, graduate students, software engineers. RB

**Theory of Computation, P.** *Lecture Notes in Computer Science-324: Mathematical Foundations of Computer Science 1988.* Ed: M.P. Chytil, L. Janiga, V. Koubek. Springer-Verlag, 1988, ix + 562 pp, $44.60 (P). [ISBN: 0-387-50110-X] Proceedings of the thirteenth such conference, held in August 1988 at Carlsbad, Czechoslovakia. An international array of authors contributed eleven invited papers and 42 short communications on a broad range of topics. RB

**Theory of Computation, S(17-18), P.** *Nonsequential Processes: A Petri Net View.* Eike Best, César Fernández C. EATCS Mono. on Theoret. Comput. Sci., V. 13. Springer-Verlag, 1988, ix + 112 pp, $39. [ISBN: 0-387-19030-9] A monograph offering a mathematical treatment of occurrence nets, which are proposed by Petri net theory for modelling concurrency by partially ordered sets. Intended for theoretical computer science students, discrete mathematicians, researchers in Petri nets, concurrent systems designers; previous knowledge of Petri nets not strictly necessary. 47 exercises. RB

**Theory of Computation, P.** *Lecture Notes in Computer Science-308: Conditional Term Rewriting*

*Systems.* Ed: S. Kaplan, J.-P. Jouannaud. Springer-Verlag, 1988, vi + 278 pp, $23.10 (P). [ISBN: 0-387-19242-5] "Conditional rewrite systems" is a subarea of term rewriting in formal language theory. Organizers of this workshop (Bordeaux, France, July 1987) hoped to identify key problems in the subarea, but unexpectedly found some presumed open questions already solved by contributed papers. Seventeen research papers, seven short descriptions of demonstrated systems. RB

**Theory of Computation, P.** *Lecture Notes in Computer Science-319: VLSI Algorithms and Architectures.* Ed: J.H. Reif. Springer-Verlag, 1988, x + 476 pp, $34.60 (P). [ISBN: 0-387-96818-0] Proceedings of AWOC 88 workshop in Corfu, Greece. Papers on parallel and NC algorithms, tree compaction, embedding of parallel networks, compaction and channel routing, VLSI layout, testing and design, distributed computing, parallel routing and sorting. RM

**Theory of Computation, S(17-18), P.** *Algebraic Theory of Processes.* Matthew Hennessy. Ser. in Found. of Comput. MIT Pr, 1988, ix + 272 pp, $39.95. [ISBN: 0-262-08171-7] A general and systematic introduction to the semantics of concurrent systems. The algebraic theory of formal semantics of languages is developed and applied to a particular semantic theory of distributed processes. Three complementary views of semantics of concurrent processes (behavioral, denotational, proof-theoretic) are introduced and shown equivalent. Self-contained for the theoretical computer scientist; about 100 exercises. RB

**Artificial Intelligence, P.** *Lecture Notes in Computer Science-313: Uncertainty and Intelligent Systems.* Ed: B. Bouchon, L. Saitta, R.R. Yager. Springer-Verlag, 1988, viii + 408 pp, $31.40 (P). [ISBN: 0-387-19402-9] Proceedings of the 2nd International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems held in Urbino, Italy, July 4-7, 1988. Contains 48 papers in the areas of evidence combination, fuzzy mathematics, fuzzy sets and possibility theory, information theory, intelligent data bases, knowledge acquisition and machine learning, knowledge representation, logics for artificial intelligence, neural nets, and reasoning techniques under uncertainty. AO

**Artificial Intelligence, P.** *Surfaces in Range Image Understanding.* Paul J. Besl. Ser. in Perception Engin. Springer-Verlag, 1988, xv + 339 pp, $58. [ISBN: 0-387-96773-7] This monograph, the author's dissertation, is a detailed account of a digital image segmentation algorithm based on 3-dimensional surface geometry, together with its theoretical motivation and experimental results. This algorithm's unified general-purpose approach does not require special domain assumptions. Extensive background material included. RB

**Artificial Intelligence, P.** *Recent Advances in Speech Understanding and Dialog Systems.* Ed: H. Niemann, M. Lang, G. Sagerer. NATO ASI Ser. F, V. 46. Springer-Verlag, 1988, x + 521 pp, $93.50. [ISBN: 0-387-19245-X] Proceedings of a two week NATO-sponsored institute in Bad Windsheim, FRG, July 1987. Ten invited surveys and 37 contributed papers: speech coding and segmentation, word recognition, linguistic processing. RB

**Artificial Intelligence, T(13-17), P, L.** *Simulation, Knowledge-Based Computing, and Fuzzy Statistics.* Constantin V. Negoita, Dan Ralescu. Elect., Comput. Sci. & Engin. Ser. Van Nostrand Reinhold, 1987, xi + 158 pp, $36.95. [ISBN: 0-442-26923-4] What are fuzzy variables and fuzzy systems, and do they have desirable qualities? The main purpose of this book is to understand these questions and provide some answers. The approach is to employ models in which the values of the variables can be linguistic rather than numeric, and casual relationships between variables can be formulated verbally rather than mathematically. The authors present a detailed description of fuzzy controllers, fuzzy statistics, and fuzzy-set theory, all basic to the study of "approximate reasoning and decision making under uncertainty." LCL

**Artificial Intelligence, S(15-17), L.** *Exploring Artificial Intelligence: Survey Talks from the National Conferences on Artificial Intelligence.* Ed: Howard E. Shrobe. Morgan Kaufmann, 1988, xii + 693 pp, $19.95 (P); $39.95. [ISBN: 0-934613-67-2; 0-934613-69-9] Revised versions of talks from the 1986-87 AAAI conferences. Gives nice overview of current artificial intelligence theory and practice. Topics include learning, natural language, planning, reasoning about causality, theoretical foundations, symbolic computing architectures and systems. Nice overview, accessible to a wide audience. RM

**Artificial Intelligence, P.** *Syntactic and Structural Pattern Recognition.* Ed: Gabriel Ferraté, *et al.* NATO ASI Ser. F, V. 45. Springer-Verlag, 1988, xv + 467 pp, $88. [ISBN: 0-387-19209-3] Proceedings of a NATO-sponsored workshop in Barcelona-Sitges, Spain, October 1986. Matching and parsing techniques in error-correcting graphs; grammatical inference under structure-simplifying constraints; integration of logic-based and syntactic techniques; applications to speech recognition, drawing recognition, feature identification, cryptosystems, medical pathology. RB

**Artificial Intelligence, P.** *Lecture Notes in Computer Science-320: Natural Language at the Computer.* Ed: A. Blaser. Springer-Verlag, 1988, 176 pp, $18 (P). [ISBN: 0-387-50011-1] Proceedings of a one-day symposium at Heidelberg, FRG (February 1988) sponsored by IBM Germany, which presented research done at IBM on natural language in text processing and human-machine communication, and forecast future prospects in the field. Seven papers on syntax and grammatical analysis, semantics, and discourse analysis. RB

**Artificial Intelligence, S(17-18), P.** *A Theory of Heuristic Information in Game-Tree Search.* Chun-Hung Tzeng. Symbolic Computat. Springer-Verlag, 1988, x + 107 pp, $39.50. [ISBN: 0-387-18665-4] A

monograph on heuristic (i.e., self-learning) computing methods which offers a theoretical foundation for the use of heuristic information, without proposing actual procedures for game-tree searches. Starting with probabilistic game models, the author formulates the notions of heuristic information and node strength, thereby allowing a mathematical theory for estimating node strength based on the heuristic information. RB

**Artificial Intelligence, P.** *Non-Standard Logics for Automated Reasoning.* Ed: Philippe Smets, *et al.* Academic Pr, 1988, x + 334 pp, $19. [ISBN: 0-12-649520-3] The publication resulting from a European project which considered various non-classical logical formalisms for describing knowledge-based systems. The book contains ten papers, each followed by brief critical discussions, which were generated by the dialogue of 27 European participants who met in workshops during September 1986 and June 1987. RM

**Artificial Intelligence, P.** *Sensors and Sensory Systems for Advanced Robots.* Ed: Paolo Dario. NATO ASI Ser. F, V. 43. Springer-Verlag, 1988, xi + 597 pp, $120. [ISBN: 0-387-19089-9] Proceedings of the NATO-sponsored workshop on the topic held at Maratea, Italy, April 1986, which gathered representatives of industry, government, and academia in biological, psychological, engineering, and computer science fields from thirteen countries. The workshop focused on analysis of transducer technologies for sensor construction; other topics are biological models, applications, integrated systems. 28 papers. RB

**Artificial Intelligence, S(15-18), L.** *Knowledge, Skill and Artificial Intelligence.* Ed: Bo Göranzon, Ingela Josefson. Foundat. & Applic. of AI. Springer-Verlag, 1988, xx + 193 pp, $39 (P). [ISBN: 0-387-19519-X] Contributions by British and Scandinavian researchers to the debate on the effects of new technology and its implications for society. Discussion of epistemological issues associated with the mechanization of tacit knowledge, issues of apprenticeship skill and artificial intelligence, artificial intelligence and the flexible craftsman. Emphasis on values, cultural and political issues relating to the introduction of artificial intelligence and expert systems. RM

**Computer Science, P.** *Data Types and Persistence.* Ed: Malcolm P. Atkinson, Peter Buneman, Ronald Morrison. Topics in Inform. Systems. Springer-Verlag, 1988, xviii + 292 pp, $49. [ISBN: 0-387-18785-5] Toward the integration of databases and programming languages, an August 1985 workshop at Appin, Scotland, focused on persistence and data types. The sixteen papers in this volume consider abstract data types, type checking of sophisticated types, binding strategies, inheritance and object oriented approaches, functional languages, concurrency, machine architectures for persistent data. RB

**Computer Science, P*.** *The Complexity of Robot Motion Planning.* John Canny. ACM Doct. Dissertat. Awards. MIT Pr, 1988, xvi + 197 pp, $27.50. [ISBN: 0-262-03136-1] The most studied and most basic robot motion planning problem is the "generalized movers problem:" find a collision-free path for a (jointed) robot in the presence of obstacles. In this truly outstanding dissertation, the author obtains exponentially faster solution algorithms by application of sophisticated mathematics (including resultants for systems of polynomials, Whitney's stratified sets from differential topology), then proceeds further to solve three long-open problems about robot motion complexity. RB

**Computer Science, P.** *Lecture Notes in Computer Science-314: Graph-Theoretic Concepts in Computer Science.* Ed: H. Göttler, H.J. Schneider. Springer-Verlag, 1988, vi + 254 pp, $23.10 (P). [ISBN: 0-387-19422-3] Proceedings of the 13th International Workshop on Graph Theoretic Concepts in Computer Science held in 1987. Papers on graph grammars and languages, layout, geometry and data structures, randomness and chaos, databases, distributed systems, program design, vision, and an application to chemistry. RM

**Computer Science, P.** *Lecture Notes in Computer Science-295: Parallel Computing in Science and Engineering.* Ed: R. Dierstein, D. Müller-Wichards, H.-M. Wacker. Springer-Verlag, 1988, 185 pp, $19.10 (P). [ISBN: 0-387-18923-8] Proceedings of a Bonn seminar (June 1987) hosted by the German Aerospace Research Establishment (DFVLR). Ten papers on background information, parallel computer architectures, software and langauges, algorithms and applications. RB

**Computer Science, P.** *Lecture Notes in Engineering-36: Japanese Supercomputing: Architecture, Algorithms, and Applications.* Ed: R.H. Mendez, S.A. Orszag. Springer-Verlag, 1988, iv + 161 pp, $27.40 (P). [ISBN: 0-387-96765-6] Keeping tabs on the Joneses: a diverse sample of twelve papers by Japanese authors related to supercomputing. Supercomputer simulations of shock waves and automobile aerodynamics; implementations of known algorithms on vector and hypercube architectures; a Japanese benchmark study of a Western superminicomputer; technical descriptions of three Japanese supercomputers. RB

**Computer Science, P.** *Lecture Notes in Computer Science-305: MFDBS 87.* Ed: J. Biskup, *et al.* Springer-Verlag, 1988, 247 pp, $24.50 (P). [ISBN: 0-387-19121-6] Thirteen papers from the First Symposium on Mathematical Fundamentals of Database Systems held January 19-23, 1987 in Dresden, GDR. AO

**Computer Science, T(16-17: 1), S, L.** *LR Parsing: Theory and Practice.* Nigel P. Chapman. Cambridge U Pr, 1987, viii + 228 pp, $39.50. [ISBN: 0-521-30413-X] Contains an introduction to the theory of LR parsing and a description of some of the practical techniques used to implement LR parsers. The theory of LR parsers is presented using parsing automata and item grammars rather than the more traditional (but less intuitive) ideas of 'valid items'

and 'valid prefixes.' AO

**Computer Science, P.** *Lecture Notes in Computer Science-293: Advances in Cryptology—CRYPTO '87.* Ed: Carl Pomerance. Springer-Verlag, 1988, x + 460 pp, $36.70 (P). [ISBN: 0-387-18796-0] Proceedings containing more than forty papers on topics such as communication networks, protocols, key distribution systems, public key systems, design, and applications. LCL

**Computer Science, P.** *Initial Computability, Algebraic Specifications, and Partial Algebras.* Horst Reichel. Internat. Ser. of Mono. on Comput. Sci., V. 2. Clarendon Pr, 1987, 221 pp, $55. [ISBN: 0-19-853806-5] Considers partial algebras and its applications to computer science. Assumes knowledge of universal algebra (especially free algebras), and data abstraction in computer science. LC

**Computer Science, P.** *Information-Based Complexity.* J.F. Traub, G.W. Wasilkowski, H. Woźniakowski. Computer Sci. & Sci. Computing. Academic Pr, 1988, xiii + 523 pp, $64.50. [ISBN: 0-12-697545-0] Information-based complexity is the branch of computational complexity theory that deals with the intrinsic difficulty of finding approximate solutions to problems for which the information available is partial, noisy, and/or only available at some (computational) cost. This book provides a comprehensive treatment of the theory and applications of information-based complexity in the worst case, average case, probabilistic, random, and asymptotic settings. AO

**Computer Science, T*(16-17: 1), L.** *Principles of Operating Systems.* Sacha Krakowiak. Transl: David Beeson. MIT Pr, 1988, xi + 469 pp, $37.50. [ISBN: 0-262-11122-5] Surveys the basic principles underlying the design of operating systems for centralized (shared memory) systems. Examples illustrating major ideas are taken from CP/M, Multics, Unix, and the operating system designed for the Plessey 250. Numerous exercises (of varying levels of difficulty) appear at the end of each chapter. AO

**Applications, T(16: 1, 2), L*.** *Wave Propagation in Solids and Fluids.* Julian L. Davis. Springer-Verlag, 1988, x + 386 pp, $75. [ISBN: 0-387-96739-7] Systematic account of the mathematical methods of wave phenomena. After a brief introduction to oscillatory phenomena and the physics of wave propagation, including the partial differential equations, chapters follow on the vibrating string, water waves, sound waves, fluid dynamics, elastic media, and the treatment of the Hamilton-Jacobi theory in the setting of variational methods. Readable. No problems. Bibliography limited to mostly well-known books and a few papers. Book to be followed by one on wave propogation in electromagnetic media. Interdisciplinary by design. Highly recommended for libraries. JK

**Applications, P.** *Computers in Mathematical Research.* Ed: N.M. Stephens, M.P. Thorne. Inst. of Math. & Its Applic, V. 14. Clarendon Pr, 1988, xii + 251 pp, $57.50. [ISBN: 0-19-853620-8] Proceedings of an IMA-sponsored conference in Cardiff, Wales (September 1986) which sampled some applications of the computer to pure and applied mathematical research, as an instrument of proof or as an exploratory tool. Sixteen papers on applications in analysis, cryptography, combinatorics, number theory, symbolic manipulation, geometry, groups, statistics, applied mathematics. RB

**Applications, S(17), P.** *Digital Signal Processing Design.* Andrew Bateman, Warren Yates. Elect. Eng., Communic., & Signal Processing. Computer Science Pr, 1989, 385 pp, $59.95. [ISBN: 0-7167-8188-3] Survey of digital signal processing technology, including algorithms and circuit design techniques, discussion of implementation of digital signal processing techniques based on TI TMS 320 series of devices. Analysis of the differences between analog and digital signal processing, discussion of application areas (filtering, spectral analysis, speech processing). Nice survey, aimed at engineers; no exercises. RM

**Applications, S(18), P.** *Ill-Posed Problems in the Natural Sciences.* Ed: A.N. Tikhonov, A.V. Goncharsky. Transl: M. Bloch. MIR (US Distr: Imported Pub), 1987, 344 pp, $9.95 (P). The title refers to "inverse" problems which arise in reconstructing physical reality from inexact experimental data. The corresponding mathematical theory treats vector operators, not necessarily linear, which may have nonunique or discontinuous inverses. Three initial papers cover general theory; the following twelve are applications to astrophysics, tomography, optics, and other areas. PZ

**Applications (Biological Science), P.** *Biomathematics and Related Computational Problems.* Ed: Luigi M. Ricciardi. Kluwer Academic, 1988, xxv + 733 pp, $149. [ISBN: 90-277-2726-0] Sixty-three papers from the International Workshop held at Naples in May 1987. Topics include nonlinear reaction-diffusion, stability of population models, and analysis of neuronal models. SP

**Applications (Biological Science), T(14-15), S, P, L.** *Quantitative Ecological Theory: An Introduction to Basic Models.* Michael R. Rose. Johns Hopkins U Pr, 1987, 203 pp, $25. [ISBN: 0-8018-3509-7] A manual designed to teach mathematical modeling in ecology to those with no prior background in modeling. It serves as an introduction to the basic models of population growth, migration, competition and coexistence, and ecosystem stability. Includes both continuous-time and discrete-time models. Assumes an exposure to calculus. LCL

**Applications (Biological Science), T(17-18: 2), P.** *Introduction to Theoretical Neurobiology, V. 1 & 2.* Henry C. Tuckwell. Stud. in Math. Biology. Cambridge U Pr, 1988, $49.50 each. *Volume 1: Linear Cable Theory and Dendritic Structure,* xii + 291 pp [ISBN: 0-521-35096-4]; *Volume 2: Nonlinear and Stochastic Theories,* xi + 265 pp. [ISBN: 0-521-35217-7] An exposition of the dynamical behavior of neurons, containing "descriptions and analyses of

the principal mathematical models that have been developed for neurons in the last 30 years." *Volume 1* covers simple neural models, derivation and time-dependent solutions of the usual linear cable equation, Rall's model nerve cell. Neuroanatomical and neurophysiological facts are reviewed and readings are suggested for those with primarily mathematical backgrounds. *Volume 2* considers more complicated and mathematically complex models, including Hodgkin-Huxley equations, stochastic models, statistical analysis of neuronal data. Intended for graduate students and researchers in applied mathematics and neurobiology. RB

**Applications (Cognitive Science), P, L*.** *Neurocomputing: Foundations of Research.* Ed: James A. Anderson, Edward Rosenfeld. MIT Pr, 1988, xxi + 729 pp, $45. [ISBN: 0-262-01097-6] Collection of 43 papers giving historical overview of the foundations of neurocomputing and connectionism. Includes early papers of McCullogh and Pitts through recent work on analog retinas by Carver Mead and colleagues. Selections from the work of Hebb, Von Neumann, Rosenblatt, Minsky and Papert, Kohonen, Grossberg, Hopfield, McClelland and Rumelhart, all give broad and balanced overview of historical development of the field. RM

**Applications (Cognitive Science), P.** *Neural Networks and Natural Intelligence.* Ed: Stephen Grossberg. MIT Pr, 1988, xi + 637 pp, $35. [ISBN: 0-262-07107-X] Collection of papers by Grossberg and colleagues at the Center for Adaptive Systems (CAS) at Boston University on perception, cognitive information processing and competitive learning, cognitive-emotional interaction, goal-oriented motor control, and robotics. The work is motivated by an in-depth analysis of biological data, computational theory following from design tradeoffs. RM

**Applications (Communication Theory), P.** *Mathematical Modelling for Information Technology: Telecommunication Transmission, Reception and Security.* Ed: A.O. Moscardini, E.H. Robson. Math. & Its Applic. Halsted Pr, 1988, 214 pp, $87.95. [ISBN: 0-470-21024-9] Proceedings of the tenth "POLYMODEL" conference held in 1987 at Sunderland Polytechnic in northeast England. Seventeen papers on signal processing, communication security, networks, and on the social impact of communication. LAS

**Applications (Communication Theory), P.** *ISDN: The Integrated Services Digital Network: Concepts, Methods, Systems.* Peter Bocker. Springer-Verlag, 1988, xi + 250 pp, $64.50. [ISBN: 0-387-17446-X] A description of ISDN with reference to its services, its technology and the operating principles of ISDN terminals, intended for implementing or operating engineers, educators, researchers. General telecommunication objectives, ISDN telecommunication services, ISDN structure, subscriber access, ISDN terminals, switching, transmission methods, user's view of ISDN. Current standards included in an appendix. RB

**Applications (Economics), S(17-18), P, L.** *Linear Programming and Economic Analysis.* Robert Dorfman, Paul A. Samuelson, Robert M. Solov. Dover, 1987, ix + 525 pp, $12.95 (P). [ISBN: 0-486-65491-5] An unabridged republication of the 1958 RAND Corporation report. Discusses the relationships betweeen linear programming and economic topics such as von Neumann game theory and Walrasian equilibrium. SM

**Applications (Engineering), P.** *Variational Principles of Continuum Mechanics with Engineering Applications, Volume 2: Introduction to Optimal Design Theory.* Vadim Komkov. Math. & Its Applic. D Reidel (US Distr: Kluwer), 1988, viii + 276 pp, $69. [ISBN: 90-277-2639-6] Follows *Volume I* (TR, June-July 1987) which reformulated the laws of continuum mechanics in variational form. This volume treats primarily two problems typified by the following example. Given a potential energy functional $V(\phi, f, \rho, \Omega)$ when $\phi$ is a stress function, $f$ the pressure, $\rho$ the mass/area, and $\Omega$ an admissible region. 1) Fix $f$ and $\Omega$ and find a $\rho$ which optimizes $V(\phi, \rho)$, and 2) fix $\rho$ and $f$ and find an $\Omega$ (a region) which optimizes $V(\phi, \Omega)$. MR

**Applications (Engineering), P.** *Design Automation: Automated Full-Custom VLSI Layout Using the ULYSSES Design Environment.* Michael L. Bushnell. Perspectives in Comput., V. 21. Academic Pr, 1988, xv + 463 pp, $44.50. [ISBN: 0-12-148400-9] System which integrates individual CAD design tools into an integrated design system that ensures proper integration and coordination of individual design tasks. Uses artificial intelligence techniques to provide an expert system for the tool integration problem, codifying a design methodology for complete VLSI layouts. RM

**Applications (Engineering), P.** *Signal Processing and Pattern Recognition in Nondestructive Evaluation of Materials.* Ed: C.H. Chen. NATO ASI Ser. F, V. 44. Springer-Verlag, 1988, viii + 344 pp, $79.50. [ISBN: 0-387-19100-3] Proceedings of a NATO-sponsored workshop held in Quebec, Canada, August 1987, which focused on applications of signal processing and artificial intelligence to the field of nondestructive evaluation and testing of materials. 22 papers, primarily on signal processing applications. RB

**Applications (Engineering), T(17-18: 1), P, L.** *Surface Engineering Geometry for Computer-Aided Design and Manufacture.* Ding Qiulin, B.J. Davies. Ser. in Mech. Engin. Halsted Pr, 1987, 340 pp, $79.95. [ISBN: 0-470-20997-6] Comprehensive treatment of surface modelling, including mathematical theory, computer solutions, systems and micro-based CAD/CAM environments, engineering applications. Includes curve and surface geometry, transformations, spline, Bezier, and rational curves and surfaces, fairing (i.e., refinement of shape quality), tool paths, surface intersection, hidden surface removal. RM

**Applications (Engineering), T*(15-16: 2-4), L.**

*Advanced Engineering Mathematics, Sixth Edition.* Erwin Kreyszig. Wiley, 1988, xviii + 1413 pp, $54.82. [ISBN: 0-471-85824-2] A book which can be used for up to four semester courses. Changes in this edition include some rewriting with a greater emphasis on readability, treatment of applications, and additional examples. Suitable for students of engineering, physics, and mathematics interested in practical problems. Modern viewpoint throughout. Self-contained. Excellent choice for courses crossing linear algebra, differential equations, complex variables, and numerical analysis. (*First Edition,* TR, December 1967; *Third Edition,* TR, October 1972 and Extended Review, June-July 1974; *Fourth Edition,* TR, June-July 1979; *Fifth Edition,* TR, February 1984.) JK

**Applications (Fluid Dynamics), P.** *Lecture Notes in Engineering-34: Multiphase Flow in Porous Media: Mechanics, Mathematics, and Numerics.* M.B. Allen III, G.A. Behie, J.A. Trangenstein. Springer-Verlag, 1988, 306 pp, $44.60 (P). [ISBN: 0-387-96731-1] Outlines what the authors emphasize to be promising new techniques in the mathematical modeling and simulation of fluid flow in petroleum resevoirs. Growing out of a series of lectures delivered in the autumn of 1986 at the IBM Scientific Center in Bergen, Norway, the monograph divides into three parts dealing in turn with the basic mechanics of oil resevoir flow (Allen); numerical analysis of resevoir fluid flow (Trangenstein); and numerical linear algebra for resevoir simulation (Behie). CE

**Applications (Physics), S(18), P.** *Mathematical Quantum Field Theory and Related Topics.* Ed: Joel S. Feldman, Lon M. Rosen. CMS Conf. Proc., V. 9. AMS, 1987, xiv + 261 pp, (P). [ISBN: 0-8218-6014-3] Proceedings of a 1987 Montreal Conference. MU

**Applications (Physics), T(17-18: 1), P.** *The Spinorial Chessboard.* P. Budinich, A. Trautman. Trieste Notes in Physics. Springer-Verlag, 1988, viii + 128 pp, $27.50 (P). [ISBN: 0-387-19078-3] A new, mathematically rigorous exposition of the theory of spinors on manifolds, via their intimate connection with Clifford algebras, with special attention paid to applications to higher-dimensional real pseudo-euclidean spaces of interest to physicists. (The title refers to a conceptual model for explaining double periodicity properties of real Clifford algebras.) RB

**Applications (Physics), S(15-18).** *Mathematical Physics.* Robert Carroll. Math. Stud., V. 152. North-Holland (US Distr: Elsevier Science), 1988, x + 399 pp, $94.75. [ISBN: 0-444-70443-4] This uninspiring text is largely devoted to proving well-known theorems, with almost no reference to physics. Contains an extensive bibliography. MU

**Applications (Physics), S(18), P.** *Mathematical Frontiers in Computational Chemical Physics.* Ed: Donald G. Truhlar. IMA, V. 15. Springer-Verlag, 1988, xii + 349 pp, $36.80. [ISBN: 0-387-96782-6] From plenary lectures given at a workshop at the University of Minnesota in 1987. Of the genre, this volume is unusually well done. Each chapter contains expository material for the nonexpert, and many unsolved problem areas are discussed. These areas include partial differential and integral equations, analytic continuation, asymptotic expansions, group theory, Lie algebra, simulations, and statistical mechanics. MU

**Applications (Physics), S(18), P.** *Gauge Field Theory and Complex Geometry.* Yuri I. Manin. Transl: N. Koblitz, J.R. King. Grund. der math. Wissenschaften, B. 289. Springer-Verlag, 1988, x + 297 pp, $79.50. [ISBN: 0-387-18275-6] This very clear translation of the 1984 original provides an excursion through the advanced introductory material necessary for the final chapter, Geometric Structures of Supersymmetry and Gravitation. MU

**Applications (Physics), S(17-18), P.** *Mathematics and General Relativity.* Ed: James A. Isenberg. Contemp. Math., V. 71. AMS, 1988, xv + 367 pp, $35 (P). [ISBN: 0-8218-5079-2] Proceedings of the AMS-IMS-SIAM Joint Summer Research Conference held in June 1986 at Santa Cruz, California. MU

**Applications (Physics), T(15: 1), L.** *A Primer of Diffusion Problems.* Richard Ghez. Wiley, 1988, xiii + 243 pp, $22.95 (P). [ISBN: 0-471-84692-9] Assumes calculus, ordinary differential equations, thermodynamics. The diffusion equation, the steady state, diffusion under external forces, time-dependent diffusion, similarity. Analytical and numerical methods for solution. Physically significant examples, mainly from metallurgy and semiconductor technology. Exercises. Concise, self-contained introduction to Laplace transforms. DFA

**Applications (Physics), S(18), P.** *Infinite Lie Algebras and Conformal Invariance in Condensed Matter and Particle Physics.* Ed: K. Dietz, V. Rittenberg. World Scientific, 1987, ix + 201 pp, $44. [ISBN: 9971-50-240-2] Proceedings of the Johns Hopkins Workshop on Current Problems in Particle Theory 10, held in Bonn, 1986. MU

**Reviewers**

RJA: Richard J. Allen, St. Olaf; DFA: David F. Appleyard, Carleton; SB: Steve Benson, St. Olaf; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; CE: Christopher Ennis, Carleton; SG: Steven Galovich, Carleton; GG: George Gilbert, St. Olaf; BH: Bruce Hanson, St. Olaf; RH: Rhonda Hatcher, St. Olaf; TH: Timothy Hesterberg, St. Olaf; PH: Paul Humke, St. Olaf; JJ: Jason Jones, St. Olaf; RBK: Roger B. Kirchner, Carleton; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; GN: Gail Nelson, Carleton; AO: Arnold Ostebee, St. Olaf; SP: Samuel Patterson, Carleton; MR: Matthew Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; GMS: G. Michael Schneider, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MS: Myriam Steinback, Macalester; ES: Elisabeth Strouse, Macalester; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; LW: Lynne Walling, St. Olaf; FLW: Frank L. Wolf, Carleton; PZ: Paul Zorn, St. Olaf.

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 96, Number 3 March 1989

# Contents

(ISSN 0002-9890)

ARTICLES

Ramanujan, Modular Equations, and Approximations to Pi or How to Compute One Billion Digits of Pi. . . . . . . J. M. BORWEIN, P. B. BORWEIN, AND D. H. BAILEY 201

Equations in Division Rings—A Survey . . . . . . . . . J. LAWRENCE AND G. SIMONS 220

LETTERS TO THE EDITOR . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 232

UNSOLVED PROBLEMS

The Missing Boundary of the Blaschke Diagram . . . . . . J. R. SANGWINE-YAGER 233

NOTES

A Geometrically Inspired Proof of the Singular Value Decomposition . . . . . . . . S. J. BLANK, NISHAN KRIKORIAN, AND DAVID SPRING 238

Sum Zero (mod $n$), Size $n$ Subsets of Integers . . . . . . . . . . . . . . . . . . . . . . . CRAIG BAILEY AND R. BRUCE RICHTER 240

On Open Maps . . . . . . . . . . . . . . . . . . . JOHN CROWE AND DOMINICK SAMPERI 242

On a Conjecture of R. J. Simpson About Exact Covering Congruences . . . . . . . . . . . . . . . . . DORON ZEILBERGER 243

THE TEACHING OF MATHEMATICS

A Note on Taylor's Theorem . . . . . . . . . . . . . . . . . JOSÉ A. FACENDA AGUIRRE 244

Material Implication Revisited . . . . . . . . . . . . . . . . . . . . . . . JOSEPH S. FULDA 247

A Pictorial Proof of Uniform Continuity . . . . . . . . . . . . . . . . . . . . D. M. BLOOM 250

On the Differentiation Formula for $\sin\theta$ . . . . . . . . . . . . . . . . . DONALD HARTIG 252

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 253

Advanced Problems and Solutions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 264

REVIEWS

Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds By G. Pólya and R. C. Read . . . . . . . . . . . . . . . . . . . . . . RUSSELL MERRIS 269

From One to Zero: A Universal History of Numbers By Georges Ifrah . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . FRANK SWETZ 273

TELEGRAPHIC REVIEWS . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 275

# NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to ANITA E. SOLOW, Department of Mathematics, Grinnell College, Grinnell, IA 50112; in geometry or topology to DENNIS DETURCK, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems (three copies) and solutions (two copies), both elementary and advanced to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, University of Calgary, Alberta, Canada T2N 1N4




Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $26 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $29. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $90 per year.





Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Assocation of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

# Ramanujan, Modular Equations, and Approximations to Pi or How to Compute One Billion Digits of Pi

J. M. BORWEIN AND P. B. BORWEIN
*Mathematics Department, Dalhousie University, Halifax, N.S. B3H 3J5 Canada*

*and*

D. H. BAILEY
*NASA Ames Research Center, Moffett Field, CA 94035*

**Preface.** The year 1987 was the centenary of Ramanujan's birth. He died in 1920 Had he not died so young, his presence in modern mathematics might be more immediately felt. Had he lived to have access to powerful algebraic manipulation software, such as MACSYMA, who knows how much more spectacular his already astonishing career might have been.

This article will follow up one small thread of Ramanujan's work which has found a modern computational context, namely, one of his approaches to approximating pi. Our experience has been that as we have come to understand these pieces of Ramanujan's work, as they have become mathematically demystified, and as we have come to realize the intrinsic complexity of these results, we have come to realize how truly singular his abilities were. This article attempts to present a considerable amount of material and, of necessity, little is presented in detail. We have, however, given much more detail than Ramanujan provided. Our intention is that the circle of ideas will become apparent and that the finer points may be pursued through the indicated references.

**1. Introduction.** There is a close and beautiful connection between the transformation theory for elliptic integrals and the very rapid approximation of pi. This connection was first made explicit by Ramanujan in his 1914 paper "Modular Equations and Approximations to $\pi$" **[26]**. We might emphasize that Algorithms 1 and 2 are not to be found in Ramanujan's work, indeed no recursive approximation of $\pi$ is considered, but as we shall see they are intimately related to his analysis. Three central examples are:

**Sum 1.** (Ramanujan)

$$\frac{1}{\pi} = \frac{\sqrt{8}}{9801} \sum_{n=0}^{\infty} \frac{(4n)!}{(n!)^4} \frac{[1103 + 26390n]}{396^{4n}}.$$

**Algorithm 1.** Let $\alpha_0 := 6 - 4\sqrt{2}$ and $y_0 := \sqrt{2} - 1$.
Let

$$y_{n+1} := \frac{1 - \left(1 - y_n^4\right)^{1/4}}{1 + \left(1 - y_n^4\right)^{1/4}}$$

and

$$\alpha_{n+1} := (1 + y_{n+1})^4 \alpha_n - 2^{2n+3} y_{n+1}\left(1 + y_{n+1} + y_{n+1}^2\right).$$

Then

$$0 < \alpha_n - 1/\pi < 16 \cdot 4^n e^{-2 \cdot 4^n \pi}$$

and $\alpha_n$ converges to $1/\pi$ *quartically* (that is, with order four).

**Algorithm 2.** Let $s_0 := 5(\sqrt{5} - 2)$ and $\alpha_0 := 1/2$.
Let

$$s_{n+1} := \frac{25}{(z + x/z + 1)^2 s_n},$$

where

$$x := 5/s_n - 1 \qquad y := (x-1)^2 + 7$$

and

$$z := \left[\frac{1}{2}x\left(y + \sqrt{y^2 - 4x^3}\right)\right]^{1/5}.$$

Let

$$\alpha_{n+1} := s_n^2 \alpha_n - 5^n \left\{ \frac{s_n^2 - 5}{2} + \sqrt{s_n\left(s_n^2 - 2s_n + 5\right)} \right\}.$$

Then

$$0 < \alpha_n - \frac{1}{\pi} < 16 \cdot 5^n e^{-5^n \pi}$$

and $\alpha_n$ converges to $1/\pi$ *quintically* (that is, with order five).

Each additional term in Sum 1 adds roughly eight digits, each additional iteration of Algorithm 1 quadruples the number of correct digits, while each additional iteration of Algorithm 2 quintuples the number of correct digits. Thus a mere thirteen iterations of Algorithm 2 provide in excess of one billion decimal digits of pi. In general, for us, $p$th-order convergence of a sequence $\{\alpha_n)$ to $\alpha$ means that $\alpha_n$ tends to $\alpha$ and that

$$|\alpha_{n+1} - \alpha| \leq C|\alpha_n - \alpha|^p$$

for some constant $C > 0$. Algorithm 1 is arguably the most efficient algorithm currently known for the extended precision calculation of pi. While the rates of convergence are impressive, it is the subtle and thoroughly nontransparent nature of these results and the beauty of the underlying mathematics that intrigue us most.

Watson **[37]**, commenting on certain formulae of Ramanujan, talks of

> *a thrill which is indistinguishable from the thrill which I feel when I enter the Sagrestia Nuovo of the Capella Medici and see before me the austere beauty of the four statues representing "Day," "Night," "Evening," and "Dawn" which Michelangelo has set over the tomb of Giuliano de'Medici and Lorenzo de'Medici.*

Sum 1 is directly due to Ramanujan and appears in **[26]**. It rests on a modular identity of order 58 and, like much of Ramanujan's work, appears without proof and with only scanty motivation. The first complete derivation we know of appears

in [11]. Algorithms 1 and 2 are based on modular identities of orders 4 and 5, respectively. The underlying quintic modular identity in Algorithm 2 (the relation for $s_n$) is also due to Ramanujan, though the first proof is due to Berndt and will appear in [7].

One intention in writing this article is to explain the genesis of Sum 1 and of Algorithms 1 and 2. It is not possible to give a short self-contained account without assuming an unusual degree of familiarity with modular function theory. Also, parts of the derivation involve considerable algebraic calculation and may most easily be done with the aid of a symbol manipulation package (MACSYMA, MAPLE, REDUCE, etc.). We hope however to give a taste of methods involved. The full details are available in [11].

A second intention is very briefly to describe the role of these and related approximations in the recent extended precision calculations of pi. In part this entails a short discussion of the complexity and implementation of such calculations. This centers on a discussion of multiplication by fast Fourier transform methods. Of considerable related interest is the fact that these algorithms for $\pi$ are provably close to the theoretical optimum.

**2. The State of Our Current Ignorance.** Pi is almost certainly the most natural of the transcendental numbers, arising as the circumference of a circle of unit diameter. Thus, it is not surprising that its properties have been studied for some twenty-five hundred years. What is surprising is how little we actually know.

We know that $\pi$ is irrational, and have known this since Lambert's proof of 1771 (see [5]). We have known that $\pi$ is transcendental since Lindemann's proof of 1882 [23]. We also know that $\pi$ is not a Liouville number. Mahler proved this in 1953. An irrational number $\beta$ is *Liouville* if, for any $n$, there exist integers $p$ and $q$ so that

$$0 < \left|\beta - \frac{p}{q}\right| < \frac{1}{q^n}.$$

Liouville showed these numbers are all transcendental. In fact we know that

$$\left|\pi - \frac{p}{q}\right| > \frac{1}{q^{14.65}} \tag{2.1}$$

for $p, q$ integral with $q$ sufficiently large. This *irrationality estimate*, due to Chudnovsky and Chudnovsky [16] is certainly not best possible. It is likely that 14.65 should be replaced by $2 + \varepsilon$ for any $\varepsilon > 0$. Almost all transcendental numbers satisfy such an inequality. We know a few related results for the rate of algebraic approximation. The results may be pursued in [4] and [11].

We know that $e^{\pi}$ is transcendental. This follows by noting that $e^{\pi} = (-1)^{-i}$ and applying the Gelfond-Schneider theorem [4]. We know that $\pi + \log 2 + \sqrt{2}\log 3$ is transcendental. This result is a consequence of the work that won Baker a Fields Medal in 1970. And we know a few more than the first two hundred million digits of the decimal expansion for $\pi$ (Kanada, see Section 3).

The state of our ignorance is more profound. We do not know whether such basic constants as $\pi + e$, $\pi/e$, or $\log \pi$ are irrational, let alone transcendental. The best we can say about these three particular constants is that they cannot satisfy any polynomial of degree eight or less with integer coefficients of average size less than $10^9$ [3]. This is a consequence of some recent computations employing the

Ferguson-Forcade algorithm [**17**]. We don't know anything of consequence about the single continued fraction of pi, except (numerically) the first 17 million terms, which Gosper computed in 1985 using Sum 1. Likewise, apart from listing the first many millions of digits of $\pi$, we know virtually nothing about the decimal expansion of $\pi$. It is possible, albeit not a good bet, that all but finitely many of the decimal digits of pi are in fact 0's and 1's. Carl Sagan's recent novel *Contact* rests on a similar possibility. Questions concerning the normality of or the distribution of digits of particular transcendentals such as $\pi$ appear completely beyond the scope of current mathematical techniques. The evidence from analysis of the first thirty million digits is that they are very uniformly distributed [**2**]. The next one hundred and seventy million digits apparently contain no surprises.

In part we perhaps settle for computing digits of $\pi$ because there is little else we can currently do. We would be amiss, however, if we did not emphasize that the extended precision calculation of pi has substantial application as a test of the "global integrity" of a supercomputer. The extended precision calculations described in Section 3 uncovered hardware errors which had to be corrected before those calculations could be successfully run. Such calculations, implemented as in Section 4, are apparently now used routinely to check supercomputers before they leave the factory. A large-scale calculation of pi is entirely unforgiving; it soaks into all parts of the machine and a single bit awry leaves detectable consequences.

## 3. Matters Computational

*I am ashamed to tell you to how many figures I carried these calculations, having no other business at the time.*

*Isaac Newton*

Newton's embarrassment at having computed 15 digits, which he did using the arcsinlike formula

$$\begin{aligned}\pi &= \frac{3\sqrt{3}}{4} + 24\left(\frac{1}{12} - \frac{1}{5\cdot 2^5} - \frac{1}{28\cdot 2^7} - \frac{1}{72\cdot 2^9} - \cdots\right)\\ &= \frac{3\sqrt{3}}{4} + 24\int_0^{\frac{1}{4}}\sqrt{x - x^2}\,dx,\end{aligned}$$

is indicative both of the spirit in which people calculate digits and the fact that a surprising number of people have succumbed to the temptation [**5**].

The history of efforts to determine an accurate value for the constant we now know as $\pi$ is almost as long as the history of civilization itself. By 2000 B.C. both the Babylonians and the Egyptians knew $\pi$ to nearly two decimal places. The Babylonians used, among others, the value 3 1/8 and the Egyptians used 3 13/81. Not all ancient societies were as accurate, however—nearly 1500 years later the Hebrews were perhaps still content to use the value 3, as the following quote suggests.

*Also, he made a molten sea of ten cubits from brim to brim, round in compass, and five cubits the height thereof; and a line of thirty cubits did compass it round about.*

*Old Testament*, 1 Kings 7:23

Despite the long pedigree of the problem, all nonempirical calculations have employed, up to minor variations, only three techniques.

i) The first technique due to Archimedes of Syracuse (287–212 B.C.) is, recursively, to calculate the length of circumscribed and inscribed regular $6 \cdot 2^n$-gons about a circle of diameter 1. Call these quantities $a_n$ and $b_n$, respectively. Then $a_0 := 2\sqrt{3}$, $b_0 := 3$ and, as Gauss's teacher Pfaff discovered in 1800,

$$a_{n+1} := \frac{2a_n b_n}{a_n + b_n} \quad \text{and} \quad b_{n+1} := \sqrt{a_{n+1} b_n}\,.$$

Archimedes, with $n = 4$, obtained

$$3\tfrac{10}{71} < \pi < 3\tfrac{1}{7}.$$

While hardly better than estimates one could get with a ruler, this is the first method that can be used to generate an arbitrary number of digits, and to a nonnumerical mathematician perhaps the problem ends here. Variations on this theme provided the basis for virtually all calculations of $\pi$ for the next 1800 years, culminating with a 34 digit calculation due to Ludolph van Ceulen (1540–1610). This demands polygons with about $2^{60}$ sides and so is extraordinarily time consuming.

ii) Calculus provides the basis for the second technique. The underlying method relies on Gregory's series of 1671

$$\arctan x = \int_0^x \frac{dt}{1+t^2} = x - \frac{x^3}{3} + \frac{x^5}{5} - \cdots \qquad |x| \le 1$$

coupled with a formula which allows small $x$ to be used, like

$$\frac{\pi}{4} = 4\arctan\left(\frac{1}{5}\right) - \arctan\left(\frac{1}{239}\right).$$

This particular formula is due to Machin and was employed by him to compute 100 digits of $\pi$ in 1706. Variations on this second theme are the basis of all the calculations done until the 1970's including William Shanks' monumental hand-calculation of 527 digits. In the introduction to his book **[32]**, which presents this calculation, Shanks writes:

> *Towards the close of the year 1850 the Author first formed the design of rectifying the circle to upwards of 300 places of decimals. He was fully aware at that time, that the accomplishment of his purpose would add little or nothing to his fame as a Mathematician though it might as a Computer; nor would it be productive of anything in the shape of pecuniary recompense.*

Shanks actually attempted to hand-calculate 707 digits but a mistake crept in at the 527th digit. This went unnoticed until 1945, when D. Ferguson, in one of the last "nondigital" calculations, computed 530 digits. Even with machine calculations mistakes occur, so most record-setting calculations are done twice—by sufficiently different methods.

The advent of computers has greatly increased the scope and decreased the toil of such calculations. Metropolis, Reitwieser, and von Neumann computed and analyzed 2037 digits using Machin's formula on ENIAC in 1949. In 1961, Dan Shanks and Wrench calculated 100,000 digits on an IBM 7090 **[31]**. By 1973, still using Machin-like arctan expansions, the million digit mark was passed by Guillard and Bouyer on a CDC 7600.

iii) The third technique, based on the transformation theory of elliptic integrals, provides the algorithms for the most recent set of computations. The most recent records are due separately to Gosper, Bailey, and Kanada. Gosper in 1985 calculated over 17 million digits (in fact over 17 million terms of the continued fraction) using a carefully orchestrated evaluation of Sum 1.

Bailey in January 1986 computed over 29 million digits using Algorithm 1 on a Cray 2 **[2]**. Kanada, using a related quadratic algorithm (due in basis to Gauss and made explicit by Brent **[12]** and Salamin **[27]**) and using Algorithm 1 for a check, verified 33,554,000 digits. This employed a HITACHI S-810/20, took roughly eight hours, and was completed in September of 1986. In January 1987 Kanada extended his computation to $2^{27}$ decimal places of $\pi$ and the hundred million digit mark had been passed. The calculation took roughly a day and a half on a NEC SX2 machine. Kanada's most recent feat (Jan. 1988) was to compute 201,326,000 digits, which required only six hours on a new Hitachi S-820 supercomputer. Within the next few years many hundreds of millions of digits will no doubt have been similarly computed. Further discussion of the history of the computation of pi may be found in **[5]** and **[9]**.

**4. Complexity Concerns.** One of the interesting morals from theoretical computer science is that many familiar algorithms are far from optimal. In order to be more precise we introduce the notion of *bit complexity*. Bit complexity counts the number of single operations required to complete an algorithm. The single-digit operations we count are $+,-,\times$. (We could, if we wished, introduce storage and logical comparison into the count. This, however, doesn't affect the order of growth of the algorithms in which we are interested.) This is a good measure of time on a serial machine. Thus, addition of two $n$-digit integers by the usual method has bit complexity $O(n)$, and straightforward uniqueness considerations show this to be asymptotically best possible.

Multiplication is a different story. Usual multiplication of two $n$-digit integers has bit complexity $O(n^2)$ and no better. However, it is possible to multiply two $n$-digit integers with complexity $O(n(\log n)(\log\log n))$. This result is due to Schönhage and Strassen and dates from 1971 **[29]**. It provides the best bound known for multiplication. No multiplication can have speed better than $O(n)$. Unhappily, more exact results aren't available.

The original observation that a faster than $O(n^2)$ multiplication is possible was due to Karatsuba in 1962. Observe that

$$(a+b10^n)(c+d10^n) = ac + [(a-b)(c-d) - ac - bd]10^n + bd10^{2n},$$

and thus multiplication of two $2n$-digit integers can be reduced to three multiplications of $n$-digit integers and a few extra additions. (Of course multiplication by $10^n$ is just a shift of the decimal point.) If one now proceeds recursively one produces a multiplication with bit complexity

$$O(n^{\log_2 3}).$$

Note that $\log_2 3 = 1.58\ldots < 2$.

We denote by $M(n)$ the bit complexity of multiplying two $n$-digit integers together by any method that is at least as fast as usual multiplication.

The trick to implementing high precision arithmetic is to get the multiplication right. Division and root extraction piggyback off multiplication using Newton's

method. One may use the iteration

$$x_{k+1} = 2x_k - x_k^2 y$$

to compute $1/y$ and the iteration

$$x_{k+1} = \frac{1}{2}\left(x_k + \frac{y}{x_k}\right)$$

to compute $\sqrt{y}$. One may also compute $1/\sqrt{y}$ from

$$x_{k+1} = \frac{x_k(3 - yx_k^2)}{2}$$

and so avoid divisions in the computation of $\sqrt{y}$. Not only do these iterations converge quadratically but, because Newton's method is self-correcting (a slight perturbation in $x_k$ does not change the limit), it is possible at the $k$th stage to work only to precision $2^k$. If division and root extraction are so implemented, they both have bit complexity $O(M(n))$, in the sense that $n$-digit input produces $n$-digit accuracy in a time bounded by a constant times the speed of multiplication. This extends in the obvious way to the solution of any algebraic equation, with the startling conclusion that every algebraic number can be computed (to $n$-digit accuracy) with bit complexity $O(M(n))$. Writing down $n$-digits of $\sqrt{2}$ or $3\sqrt{7}$ is (up to a constant) no more complicated than multiplication.

The Schönhage-Strassen multiplication is hard to implement. However, a multiplication with complexity $O((\log n)^{2+\varepsilon} n)$ based on an ordinary complex (floating point) fast Fourier transform is reasonably straightforward. This is Kanada's approach, and the recent records all rely critically on some variations of this technique.

To see how the fast Fourier transform may be used to accelerate multiplication, let $x := (x_0, x_1, x_2, \ldots, x_{n-1})$ and $y := (y_0, y_1, y_2, \ldots, y_{n-1})$ be the representations of two high-precision numbers in some radix $b$. The radix $b$ is usually selected to be some power of 2 or 10 whose square is less than the largest integer exactly representable as an ordinary floating-point number on the computer being used. Then, except for releasing each "carry," the product $z := (z_0, z_1, z_2, \ldots, z_{2n-1})$ of $x$ and $y$ may be written as

$$\begin{aligned}
z_0 &= x_0 y_0 \\
z_1 &= x_0 y_1 + x_1 y_0 \\
z_2 &= x_0 y_2 + x_1 y_1 + x_2 y_0 \\
&\vdots \\
z_{n-1} &= x_0 y_{n-1} + x_1 y_{n-2} + \cdots + x_{n-1} y_0 \\
&\vdots \\
z_{2n-3} &= x_{n-1} y_{n-2} + x_{n-2} y_{n-1} \\
z_{2n-2} &= x_{n-1} y_{n-1} \\
z_{2n-1} &= 0.
\end{aligned}$$

Now consider $x$ and $y$ to have $n$ zeros appended, so that $x$, $y$, and $z$ all have length $N = 2n$. Then a key observation may be made: the product sequence $z$ is

precisely the discrete convolution $C(x, y)$:

$$z_k = C_k(x, y) = \sum_{j=0}^{N-1} x_j y_{k-j},$$

where the subscript $k - j$ is to be interpreted as $k - j + N$ if $k - j$ is negative.

Now a well-known result of Fourier analysis may be applied. Let $F(x)$ denote the *discrete Fourier transform* of the sequence $x$, and let $F^{-1}(x)$ denote the inverse discrete Fourier transform of $x$:

$$F_k(x) := \sum_{j=0}^{N-1} x_j e^{-2\pi ijk/N}$$

$$F_k^{-1}(x) := \frac{1}{N} \sum_{j=0}^{N-1} x_j e^{2\pi ijk/N}.$$

Then the "convolution theorem," whose proof is a straightforward exercise, states that

$$F[C(x, y)] = F(x)F(y)$$

or, expressed another way,

$$C(x, y) = F^{-1}[F(x)F(y)].$$

Thus the entire multiplication pyramid $z$ can be obtained by performing two forward discrete Fourier transforms, one vector complex multiplication and one inverse transform, each of length $N = 2n$. Once the real parts of the resulting complex numbers have been rounded to the nearest integer, the final multiprecision product may be obtained by merely releasing the carries modulo $b$. This may be done by starting at the end of the $z$ vector and working backward, as in elementary school arithmetic, or by applying other schemes suitable for vector processing on more sophisticated computers.

A straightforward implementation of the above procedure would not result in any computational savings—in fact, it would be several times more costly than the usual "schoolboy" scheme. The reason this scheme is used is that the discrete Fourier transform may be computed much more rapidly using some variation of the well-known "fast Fourier transform" (FFT) algorithm [**13**]. In particular, if $N = 2^m$, then the discrete Fourier transform may be evaluated in only $5m2^m$ arithmetic operations using an FFT. Direct application of the definition of the discrete Fourier transform would require $2^{2m+3}$ floating-point arithmetic operations, even if it is assumed that all powers of $e^{-2\pi i/N}$ have been precalculated.

This is the basic scheme for high-speed multiprecision multiplication. Many details of efficient implementations have been omitted. For example, it is possible to take advantage of the fact that the input sequences $x$ and $y$ and the output sequence $z$ are all purely real numbers, and thereby sharply reduce the operation count. Also, it is possible to dispense with complex numbers altogether in favor of performing computations in fields of integers modulo large prime numbers. Interested readers are referred to [**2**], [**8**], [**13**], and [**22**].

When the costs of all the constituent operations, using the best known techniques, are totalled both Algorithms 1 and 2 compute $n$ digits of $\pi$ with bit complexity $O(M(n)\log n)$, and use $O(\log n)$ full precision operations.

The bit complexity for Sum 1, or for $\pi$ using any of the arctan expansions, is between $O((\log n)^2 M(n))$ and $O(nM(n))$ depending on implementation. In each case, one is required to sum $O(n)$ terms of the appropriate series. Done naively, one obtains the latter bound. If the calculation is carefully orchestrated so that the terms are grouped to grow evenly in size (as rational numbers) then one can achieve the former bound, but with no corresponding reduction in the number of operations.

The Archimedean iteration of section 2 converges like $1/4^n$ so in excess of $n$ iterations are needed for $n$-digit accuracy, and the bit complexity is $O(nM(n))$.

Almost any familiar transcendental number such as $e$, $\gamma$, $\zeta(3)$, or Catalan's constant (presuming the last three to be nonalgebraic) can be computed with bit complexity $O((\log n)M(n))$ or $O((\log n)^2 M(n))$. None of these numbers is known to be computable essentially any faster than this. In light of the previous observation that algebraic numbers are all computable with bit complexity $O(M(n))$, a proof that $\pi$ cannot be computed with this speed would imply the transcendence of $\pi$. It would, in fact, imply more, as there are transcendental numbers which have complexity $O(M(n))$. An example is $0.10100100001\ldots$ .

It is also reasonable to speculate that computing the $n$th digit of $\pi$ is not very much easier than computing all the first $n$ digits. We think it very probable that computing the $n$th digit of $\pi$ cannot be $O(n)$.

## 5. The Miracle of Theta Functions

*When I was a student, abelian functions were, as an effect of the Jacobian tradition, considered the uncontested summit of mathematics, and each of us was ambitious to make progress in this field. And now? The younger generation hardly knows abelian functions.*

*Felix Klein* [**21**]

Felix Klein's lament from a hundred years ago has an uncomfortable timelessness to it. Sadly, it is now possible never to see what Bochner referred to as "the miracle of the theta functions" in an entire university mathematics program. A small piece of this miracle is required here [**6**], [**11**], [**28**]. First some standard notations. The *complete elliptic integrals of the first and second kind*, respectively,

$$K(k) := \int_0^{\pi k} \frac{dt}{\sqrt{1 - k^2 \sin^2 t}} \tag{5.1}$$

and

$$E(k) := \int_0^{\pi/2} \sqrt{1 - k^2 \sin^2 t}\, dt. \tag{5.2}$$

The second integral arises in the rectification of the ellipse, hence the name elliptic integrals. The *complementary modulus* is

$$k' := \sqrt{1 - k^2}$$

and the *complementary integrals* $K'$ and $E'$ are defined by

$$K'(k) := K(k') \quad \text{and} \quad E'(k) := E(k').$$

The first remarkable identity is *Legendre's relation* namely

$$E(k)K'(k) + E'(k)K(k) - K(k)K'(k) = \frac{\pi}{2} \tag{5.3}$$

(for $0 < k < 1$), which is pivotal in relating these quantities to pi. We also need to define two *Jacobian theta functions*

$$\Theta_2(q) := \sum_{n=-\infty}^{\infty} q^{(n+1/2)^2} \tag{5.4}$$

and

$$\Theta_3(q) := \sum_{n=-\infty}^{\infty} q^{n^2}. \tag{5.5}$$

These are in fact specializations with ($t = 0$) of the general theta functions. More generally

$$\Theta_3(t, q) := \sum_{n=-\infty}^{\infty} q^{n^2} e^{\operatorname{im} t} \qquad (\operatorname{im} t > 0)$$

with similar extensions of $\Theta_2$. In Jacobi's approach these general theta functions provide the basic building blocks for elliptic functions, as functions of $t$ (see **[11]**, **[39]**).

The complete elliptic integrals and the special theta functions are related as follows. For $|q| < 1$

$$K(k) = \frac{\pi}{2}\Theta_3^2(q) \tag{5.6}$$

and

$$E(k) = (k')^2\left[K(k) + k\frac{dK(k)}{dk}\right], \tag{5.7}$$

where

$$k := k(q) = \frac{\Theta_2^2(q)}{\Theta_3^2(q)}, \qquad k' := k'(q) = \frac{\Theta_3^2(-q)}{\Theta_3^2(q)} \tag{5.8}$$

and

$$q = e^{-\pi K'(k)/K(k)}. \tag{5.9}$$

The *modular function* $\lambda$ is defined by

$$\lambda(t) := \lambda(q) := k^2(q) := \left[\frac{\Theta_2(q)}{\Theta_3(q)}\right]^4, \tag{5.10}$$

where

$$q := e^{i\pi t}.$$

We wish to make a few comments about modular functions in general before restricting our attention to the particular modular function $\lambda$. *Modular functions* are functions which are meromorphic in $H$, the upper half of the complex plane, and which are invariant under a group of linear fractional transformations, $G$, in the sense that

$$f(g(z)) = f(z) \qquad \forall g \in G.$$

[Additional growth conditions on $f$ at certain points of the associated fundamental region (see below) are also demanded.] We restrict $G$ to be a subgroup of the *modular group* $\Gamma$ where $\Gamma$ is the set of all transformations $w$ of the form

$$w(t) = \frac{at+b}{ct+d},$$

with $a, b, c, d$ integers and $ad - bc = 1$. Observe that $\Gamma$ is a group under composition. A *fundamental region* $F_G$ is a set in $H$ with the property that any element in $H$ is uniquely the image of some element in $F_G$ under the action of $G$. Thus the behaviour of a modular function is uniquely determined by its behaviour on a fundamental region.

Modular functions are, in a sense, an extension of elliptic (or doubly periodic) functions—functions such as *sn* which are invariant under linear transformations and which arise naturally in the inversion of elliptic integrals.

The definitions we have given above are not complete. We will be more precise in our discussion of $\lambda$. One might bear in mind that much of the theory for $\lambda$ holds in considerably greater generality.

The *fundamental region* $F$ we associate with $\lambda$ is the set of complex numbers

$$F := \{\operatorname{im} t \geq 0\} \cap \big[\{|\operatorname{re} t| < 1 \quad \text{and} \\ |2t \pm 1| > 1\} \cup \{\operatorname{re} t = -1\} \cup \{|2t+1| = 1\}\big].$$

The $\lambda$-*group* (or theta-subgroup) is the set of linear fractional transformations $w$ satisfying

$$w(t) := \frac{at+b}{ct+d},$$

where $a, b, c, d$ are integers and $ad - bc = 1$, while in addition $a$ and $d$ are odd and $b$ and $c$ are even. Thus the corresponding matrices are unimodular. What makes $\lambda$ a $\lambda$-modular function is the fact that $\lambda$ is meromorphic in $\{\operatorname{im} t > 0\}$ and that

$$\lambda(w(t)) := \lambda(t)$$

for all $w$ in the $\lambda$-group, plus the fact that $\lambda$ tends to a definite limit (possibly infinite) as $t$ tends to a vertex of the fundamental region (one of the three points $(0, -1)$, $(0, 0)$, $(i, \infty)$). Here we only allow convergence from within the fundamental region.

Now some of the miracle of modular functions can be described. Largely because every point in the upper half plane is the image of a point in $F$ under an element of the $\lambda$-group, one can deduce that any $\lambda$-modular function that is bounded on $F$ is constant. Slightly further into the theory, but relying on the above, is the result that any two modular functions are algebraically related, and resting on this, but further again into the field, is the following remarkable result. Recall that $q$ is given by (5.9).

THEOREM 1. *Let $z$ be a primitive pth root of unity for $p$ an odd prime. Consider the pth order modular equation for $\lambda$ as defined by*

$$W_p(x, \lambda) := (x - \lambda_0)(x - \lambda_1) \cdots (x - \lambda_p), \tag{5.11}$$

*where*

$$\lambda_i := \lambda(z^i q^{1/p}) \qquad i < p$$

*and*

$$\lambda_p := \lambda(q^p).$$

*Then the function $W_p$ is a polynomial in $x$ and $\lambda$ (**independent** of $q$), which has integer coefficients and is of degree $p + 1$ in both $x$ and $\lambda$.*

The modular equation for $\lambda$ usually has a simpler form in the associated variables $u := x^{1/8}$ and $v := \lambda^{1/8}$. In this form the 5th-order modular equation is given by

$$\Omega_5(u, v) := u^6 - v^6 + 5u^2v^2(u^2 - v^2) + 4uv(1 - u^4v^4). \tag{5.12}$$

In particular

$$\frac{\Theta_2(q^p)}{\Theta_3(q^p)} = v^2 \quad \text{and} \quad \frac{\Theta_2(q)}{\Theta_3(q)} = u^2$$

are related by an algebraic equation of degree $p + 1$.

The miracle is not over. The *pth-order multiplier* (for $\lambda$) is defined by

$$M_p(k(q), k(q^p)) := \frac{K(k(q^p))}{K(k(q))} = \left[\frac{\Theta_3(q^p)}{\Theta_3(q)}\right]^2 \tag{5.13}$$

and turns out to be a rational function of $k(q^p)$ and $k(q)$.

One is now in possession of a $p$th-order algorithm for $K/\pi$, namely: Let $k_i := k(q^{p^i})$. Then

$$\frac{2K(k_0)}{\pi} = M_p^{-1}(k_0, k_1)M_p^{-1}(k_1, k_2)M_p^{-1}(k_2, k_3)\cdots.$$

This is an entirely algebraic algorithm. One needs to know the $p$th-order modular equation for $\lambda$ to compute $k_{i+1}$ from $k_i$ and one needs to know the rational multiplier $M_p$. The speed of convergence ($O(c^{p^i})$, for some $c < 1$) is easily deduced from (5.13) and (5.9).

The function $\lambda(t)$ is 1-1 on $F$ and has a well-defined inverse, $\lambda^{-1}$, with branch points only at 0, 1 and $\infty$. This can be used to provide a one line proof of the "big" Picard theorem that a nonconstant entire function misses at most one value (as does exp). Indeed, suppose $g$ is an entire function and that it is never zero or one; then $\exp(\lambda^{-1}(g(z)))$ is a bounded entire function and is hence constant.

Littlewood suggested that, at the right point in history, the above would have been a strong candidate for a 'one line doctoral thesis'.

**6. Ramanujan's Solvable Modular Equations.** Hardy **[19]** commenting on Ramanujan's work on elliptic and modular functions says

> *It is here that both the profundity and limitations of Ramanujan's knowledge stand out most sharply.*

We present only one of Ramanujan's modular equations.

THEOREM 2.

$$\frac{5\Theta_3(q^{25})}{\Theta_3(q)} = 1 + r_1^{1/5} + r_2^{1/5}, \tag{6.1}$$

*where for* $i = 1$ *and* 2

$$r_1 := \tfrac{1}{2}x\left(y \pm \sqrt{y^2 - 4x^3}\right)$$

*with*

$$x := \frac{5\Theta_3(q^5)}{\Theta_3(q)} - 1 \quad \text{and} \quad y := (x-1)^2 + 7.$$

This is a slightly rewritten form of entry 12(iii) of Chapter 19 of Ramanujan's *Second Notebook* (see [7], where Berndt's proofs may be studied). One can think of Ramanujan's quintic modular equation as an equation in the multiplier $M_p$ of (5.13). The initial surprise is that it is solvable. The quintic modular relation for $\lambda$, $W_5$, and the related equation for $\lambda^{1/8}$, $\Omega_5$, are both nonsolvable. The Galois group of the sixth-degree equation $\Omega_5$ (see (5.12)) over $\mathbb{Q}(v)$ is $A_5$ and is nonsolvable. Indeed both Hermite and Kronecker showed, in the middle of the last century, that the solution of a general quintic may be effected in terms of the solution of the 5th-order modular equation (5.12) and the roots may thus be given in terms of the theta functions.

In fact, in general, the Galois group for $W_p$ of (5.11) has order $p(p+1)(p-1)$ and is never solvable for $p \geq 5$. The group is quite easy to compute, it is generated by two permutations. If

$$q := e^{i\pi t}, \quad \text{then} \quad \tau \to \tau + 2 \quad \text{and} \quad \tau \to \frac{\tau}{(2\tau+1)}$$

are both elements of the $\lambda$-group and induce permutations on the $\lambda_i$ of Theorem 1. For any fixed $p$, one can use the $q$-expansion of (5.10) to compute the effect of these transformations on the $\lambda_i$, and can thus easily write down the Galois group.

While $W_p$ is not solvable over $\mathbb{Q}(\lambda)$, it is solvable over $\mathbb{Q}(\lambda, \lambda_0)$. Note that $\lambda_0$ is a root of $W_p$. It is of degree $p+1$ because $W_p$ is irreducible. Thus the Galois group for $W_p$ over $\mathbb{Q}(\lambda, \lambda_0)$ has order $p(p-1)$. For $p = 5$, 7, and 11 this gives groups of order 20, 42, and 110, respectively, which are obviously solvable and, in fact, for general primes, the construction always produces a solvable group.

From (5.8) and (5.10) one sees that Ramanujan's modular equation can be rewritten to give $\lambda_5$ solvable in terms of $\lambda_0$ and $\lambda$. Thus, we can hope to find an explicit solvable relation for $\lambda_p$ in terms of $\lambda$ and $\lambda_0$. For $p = 3$, $W_p$ is of degree 4 and is, of course, solvable. For $p = 7$, Ramanujan again helps us out, by providing a solvable seventh-order modular identity for the closely related *eta function* defined by

$$\eta(q) := q^{\frac{1}{12}} \prod_{n=1}^{\infty} (1 - q^{2n}).$$

The first interesting prime for which an explicit solvable form is not known is the "endecadic" ($p = 11$) case. We consider only prime values because for nonprime values the modular equation factors.

This leads to the interesting problem of mechanically constructing these equations. In principle, and to some extent in practice, this is a purely computational problem. Modular equations can be computed fairly easily from (5.11) and even more easily in the associated variables $u$ and $v$. Because one knows a priori bounds on the size of the (integer) coefficients of the equations one can perform these calculations exactly. The coefficients of the equation, in the variables $u$ and $v$, grow at most like $2^n$. (See **[11]**.) Computing the solvable forms and the associated computational problems are a little more intricate—though still in principle entirely mechanical. A word of caution however: in the variables $u$ and $v$ the endecadic modular equation has largest coefficient 165, a three digit integer. The endecadic modular equation for the intimately related function $J$ (Klein's *absolute invariant*) has coefficients as large as

$$27090964785531389931563200281035226311929052227303 \times 2^{92}3^{19}5^{20}11^{2}53.$$

It is, therefore, one thing to solve these equations, it is entirely another matter to present them with the economy of Ramanujan.

The paucity of Ramanujan's background in complex analysis and group theory leaves open to speculation Ramanujan's methods. The proofs given by Berndt are difficult. In the seventh-order case, Berndt was aided by MACSYMA—a sophisticated algebraic manipulation package. Berndt comments after giving the proof of various seventh-order modular identities:

> *Of course, the proof that we have given is quite unsatisfactory because it is a verification that could not have been achieved without knowledge of the result. Ramanujan obviously possessed a more natural, transparent, and ingenious proof.*

**7. Modular Equations and Pi.** We wish to connect the modular equations of Theorem 1 to pi. This we contrive via the function *alpha* defined by:

$$\alpha(r) := \frac{E'(k)}{K(k)} - \frac{\pi}{(2K(k))^2}, \tag{7.1}$$

where

$$k := k(q) \quad \text{and} \quad q := e^{-\pi\sqrt{r}}.$$

This allows one to rewrite Legendre's equation (5.3) in a one-sided form without the conjugate variable as

$$\frac{\pi}{4} = K\left[\sqrt{r}E - \left(\sqrt{r} - \alpha(r)\right)K\right]. \tag{7.2}$$

We have suppressed, and will continue to suppress, the $k$ variable. With (5.6) and (5.7) at hand we can write a $q$-expansion for $\alpha$, namely,

$$\alpha(r) = \frac{\dfrac{1}{\pi} - \sqrt{r}\,4\dfrac{\sum_{n=-\infty}^{\infty} n^2(-q)^{n^2}}{\sum_{n=-\infty}^{\infty} (-q)^{n^2}}}{\left[\sum_{n=-\infty}^{\infty} q^{n^2}\right]^4}, \tag{7.3}$$

and we can see that as $r$ tends to infinity $q = e^{-\pi\sqrt{r}}$ tends to zero and $\alpha(r)$ tends to $1/\pi$. In fact

$$\alpha(r) - \frac{1}{\pi} \approx 8\left(\sqrt{r} - \frac{1}{\pi}\right)e^{-\pi\sqrt{r}}. \tag{7.4}$$

The key now is iteratively to calculate $\alpha$. This is the content of the next theorem.

THEOREM 3. *Let $k_0 := k(q)$, $k_1 := k(q^p)$ and $M_p := M_p(k_0, k_1)$ as in* (5.13). *Then*

$$\alpha(p^2 r) = \frac{\alpha(r)}{M_p^2} - \sqrt{r}\left[\frac{k_0^2}{M_p^2} - pk_1^2 + \frac{pk_1'^2 k_1 \dot{M}_p}{M_p}\right],$$

*where represents the full derivative of $M_p$ with respect to $k_0$. In particular, $\alpha$ is algebraic for rational arguments.*

We know that $K(k_1)$ is related via $M_p$ to $K(k)$ and we know that $E(k)$ is related via differentiation to $K$. (See (5.7) and (5.13).) Note that $q \to q^p$ corresponds to $r \to p^2 r$. Thus from (7.2) some relation like that of the above theorem must exist. The actual derivation requires some careful algebraic manipulation. (See **[11]**, where it has also been made entirely explicit for $p := 2$, 3, 4, 5, and 7, and where numerous algebraic values are determined for $\alpha(r)$.) In the case $p := 5$ we can specialize with some considerable knowledge of quintic modular equations to get:

THEOREM 4. *Let $s := 1/M_5(k_0, k_1)$. Then*

$$\alpha(25r) = s^2\alpha(r) - \sqrt{r}\left[\frac{(s^2-5)}{2} + \sqrt{s(s^2-2s+5)}\right].$$

This couples with Ramanujan's quintic modular equation to provide a derivation of Algorithm 2.

Algorithm 1 results from specializing Theorem 3 with $p := 4$ and coupling it with a quartic modular equation. The quartic equation in question is just two steps of the corresponding quadratic equation which is Legendre's form of the *arithmetic geometric mean iteration*, namely:

$$k_1 = \frac{2\sqrt{k}}{1+k}.$$

An algebraic $p$th-order algorithm for $\pi$ is derived from coupling Theorem 3 with a $p$th-order modular equation. The substantial details which are skirted here are available in **[11]**.

## 8. Ramanujan's sum.

**8. Ramanujan's sum.** This amazing sum,

$$\frac{1}{\pi} = \frac{\sqrt{8}}{9801}\sum_{n=0}^{\infty}\frac{(4n)!}{(n!)^4}\frac{[1103 + 26390n]}{396^{4n}}$$

is a specialization ($N = 58$) of the following result, which gives reciprocal series for $\pi$ in terms of our function alpha and related modular quantities.

THEOREM 5.

$$\frac{1}{\pi} = \sum_{n=0}^{\infty} \frac{\left(\frac{1}{4}\right)_n \left(\frac{1}{2}\right)_n \left(\frac{3}{4}\right)_n d_n(N)}{(n!)^3} x_N^{2n+1}, \tag{8.1}$$

*where,*

$$x_N := \frac{4k_N(k_N')^2}{\left(1+k_N^2\right)^2} := \left(\frac{g_N^{12}+g_N^{-12}}{2}\right)^{-1},$$

*with*

$$d_n(N) = \left[\frac{\alpha(N)x_N^{-1}}{1+k_N^2} - \frac{\sqrt{N}}{4}g_N^{-12}\right] + n\sqrt{N}\left(\frac{g_N^{12}-g_N^{-12}}{2}\right)$$

*and*

$$k_N := k\left(e^{-\pi\sqrt{N}}\right), \qquad g_N^{12} = (k_N')^2/(2k_N).$$

*Here $(c)_n$ is the rising factorial: $(c)_n := c(c+1)(c+2)\cdots(c+n-1)$.*

Some of the ingredients for the proof of Theorem 5, which are detailed in [**11**], are the following. Our first step is to write (7.2) as a sum after replacing the $E$ by $K$ and $dK/dk$ using (5.7). One then uses an identity of Clausen's which allows one to write the square of a hypergeometric function ${}_2F_1$ in terms of a generalized hypergeometric ${}_3F_2$, namely, for all $k$ one has

$$(1+k^2)\left[\frac{2K(k)}{\pi}\right]^2 = {}_3F_2\left(\frac{1}{4},\frac{3}{4},\frac{1}{2},1,1:\left(\frac{2}{g^{12}+g^{-12}}\right)^2\right)$$
$$= \sum_{n=0}^{\infty} \frac{\left(\frac{1}{4}\right)_n\left(\frac{3}{4}\right)_n\left(\frac{1}{2}\right)_n}{(1)_n(1)_n} \frac{\left(\frac{2}{g^{12}+g^{-12}}\right)^{2n}}{n!}.$$

Here g is related to $k$ by

$$\frac{4k(k')^2}{(1+k^2)^2} = \left(\frac{g^{12}+g^{-12}}{2}\right)^{-1}$$

as required in Theorem 5. We have actually done more than just use Clausen's identity, we have also transformed it once using a standard hypergeometric substitution due to Kummer. Incidentally, Clausen was a nineteenth-century mathematician who, among other things, computed 250 digits of $\pi$ in 1847 using Machin's formula. The desired formula (8.1) is obtained on combining these pieces.

Even with Theorem 5, our work is not complete. We still have to compute

$$k_{58} := k\left(e^{-\pi\sqrt{58}}\right) \quad \text{and} \quad \alpha_{58} := \alpha(58).$$

In fact

$$g_{58}^2 = \left(\frac{\sqrt{29}+5}{2}\right)$$

is a well-known *invariant* related to the fundamental solution to Pell's equation for 29 and it turns out that

$$\alpha_{58} = \left(\frac{\sqrt{29}+5}{2}\right)^6 (99\sqrt{29} - 444)(99\sqrt{2} - 70 - 13\sqrt{29}).$$

One can, in principle, and for $N := 58$, probably in practice, solve for $k_N$ by directly solving the $N$th-order equation

$$W_N\left(k_N^2, 1 - k_N^2\right) = 0.$$

For $N = 58$, given that Ramanujan **[26]** and Weber **[38]** have calculated $g_{58}$ for us, verification by this method is somewhat easier though it still requires a tractable form of $W_{58}$. Actually, more sophisticated number-theoretic techniques exist for computing $k_N$ (these numbers are called *singular moduli*). A description of such techniques, including a reconstruction of how Ramanujan might have computed the various singular moduli he presents in **[26]**; is presented by Watson in a long series of papers commencing with **[36]**; and some more recent derivations are given in **[11]** and **[30]**. An inspection of Theorem 5 shows that all the constants in Series 1 are determined from $g_{58}$. Knowing $\alpha$ is equivalent to determining that the number 1103 is correct.

It is less clear how one explicitly calculates $\alpha_{58}$ in algebraic form, except by brute force, and a considerable amount of brute force is required; but a numerical calculation to any reasonable accuracy is easily obtained from (7.3) and 1103 appears! The reader is encouraged to try this to, say, 16 digits. This presumably is what Ramanujan observed. Ironically, when Gosper computed 17 million digits of $\pi$ using Sum 1, he had no mathematical proof that Sum 1 actually converged to $1/\pi$. He compared ten million digits of the calculation to a previous calculation of Kanada et al. This verification that Sum 1 is correct to ten million places also provided the first complete proof that $\alpha_{58}$ is as advertised above. A nice touch—that the calculation of the sum should prove itself as it goes.

Roughly this works as follows. One knows enough about the exact algebraic nature of the components of $d_n(N)$ and $x_N$ to know that if the purported sum (of positive terms) were incorrect, that before one reached 3 million digits, this sum must have ceased to agree with $1/\pi$. Notice that the components of Sum 1 are related to the solution of an equation of degree 58, but virtually no irrationality remains in the final packaging. Once again, there are very good number-theoretic reasons, presumably unknown to Ramanujan, why this must be so (58 is at least a good candidate number for such a reduction). Ramanujan's insight into this marvellous simplification remains obscure.

Ramanujan **[26]** gives 14 other series for $1/\pi$, some others almost as spectacular as Sum 1—and one can indeed derive some even more spectacular related series.* He gives almost no explanation as to their genesis, saying only that there are "corresponding theories" to the standard theory (as sketched in section 5) from which they follow. Hardy, quoting Mordell, observed that "it is unfortunate that Ramanujan has not developed the corresponding theories." By methods analogous

*(Added in proof) Many related series due to Borwein and Borwein and to Chudnovsky and Chudnovsky appear in papers in *Ramanujan Revisited*, Academic Press, 1988.

to those used above, all his series can be derived from the classical theory **[11]**. Again it is unclear what passage Ramanujan took to them, but it must in some part have diverged from ours.

We conclude by writing down another extraordinary series of Ramanujan's, which also derives from the same general body of theory,

$$\frac{1}{\pi} = \sum_{n=0}^{\infty} \binom{2n}{n}^3 \frac{42n+5}{2^{12n+4}}.$$

This series is composed of fractions whose numerators grow like $2^{6n}$ and whose denominators are exactly $16 \cdot 2^{12n}$. In particular this can be used to calculate the second block of $n$ binary digits of $\pi$ without calculating the first $n$ binary digits. This beautiful observation, due to Holloway, results, disappointingly, in no intrinsic reduction in complexity.

**9. Sources.** References **[7]**, **[11]**, **[19]**, **[26]**, **[36]**, and **[37]** relate directly to Ramanujan's work. References **[2]**, **[8]**, **[9]**, **[10]**, **[12]**, **[22]**, **[24]**, **[27]**, **[29]**, and **[31]** discuss the computational concerns of the paper.

Material on modular functions and special functions may be pursued in **[1]**, **[6]**, **[9]**, **[14]**, **[15]**, **[18]**, **[20]**, **[28]**, **[34]**, **[38]**, and **[39]**. Some of the number-theoretic concerns are touched on in **[3]**, **[6]**, **[9]**, **[11]**, **[16]**, **[23]**, and **[35]**.

Finally, details of all derivations are given in **[11]**.

REFERENCES

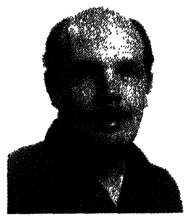
1. M. Abramowitz and I. Stegun, Handbook of Mathematical Functions, Dover, New York, 1964.
2. D. H. Bailey, The Computation of $\pi$ to 29,360,000 decimal digits using Borweins' quartically convergent algorithm, *Math. Comput.*, 50 (1988) 283–96.
3. ______, Numerical results on the transcendence of constants involving $\pi$, $e$, and Euler's constant, *Math. Comput.*, 50 (1988) 275–81.
4. A. Baker, Transcendental Number Theory, Cambridge Univ. Press, London, 1975.
5. P. Beckmann, A History of Pi, 4th ed., Golem Press, Boulder, CO, 1977.
6. R. Bellman, A Brief Introduction to Theta Functions, Holt, Reinhart and Winston, New York, 1961.
7. B. C. Berndt, Modular Equations of Degrees 3, 5, and 7 and Associated Theta Functions Identities, chapter 19 of Ramanujan's Second Notebook, Springer—to be published.
8. A. Borodin and I. Munro, The Computational Complexity of Algebraic and Numeric Problems, American Elsevier, New York, 1975.
9. J. M. Borwein and P. B. Borwein, The arithmetic-geometric mean and fast computation of elementary functions, *SIAM Rev.*, 26 (1984), 351–365.
10. ______, An explicit cubic iteration for pi, *BIT*, 26 (1986) 123–126.
11. ______, Pi and the AGM—A Study in Analytic Number Theory and Computational Complexity, Wiley, N.Y., 1987.
12. R. P. Brent, Fast multiple-precision evaluation of elementary functions, *J. ACM*, 23 (1976) 242–251.
13. E. O. Brigham, The Fast Fourier Transform, Prentice-Hall, Englewood Cliffs, N.J., 1974.
14. A. Cayley, An Elementary Treatise on Elliptic Functions, Bell and Sons, 1885; reprint Dover, 1961.
15. A. Cayley, A memoir on the transformation of elliptic functions, *Phil. Trans. T.*, 164 (1874) 397–456.
16. D. V. Chudnovsky and G. V. Chudnovsky, Padé and Rational Approximation to Systems of Functions and Their Arithmetic Applications, Lecture Notes in Mathematics 1052, Springer, Berlin, 1984.
17. H. R. P. Ferguson and R. W. Forcade, Generalization of the Euclidean algorithm for real numbers to all dimensions higher than two, *Bull. AMS*, 1 (1979) 912–914.
18. C. F. Gauss, Werke, Göttingen 1866–1933, Bd 3, pp. 361–403.

19. G. H. Hardy, Ramanujan, Cambridge Univ. Press, London, 1940.
20. L. V. King, On The Direct Numerical Calculation of Elliptic Functions and Integrals, Cambridge Univ. Press, 1924.
21. F. Klein, Development of Mathematics in the 19th Century, 1928, Trans Math Sci. Press, R. Hermann ed., Brookline, MA, 1979.
22. D. Knuth, The Art of Computer Programming, vol. 2: Seminumerical Algorithms, Addison-Wesley, Reading, MA, 1981.
23. F. Lindemann, Über die Zahl $\pi$, *Math. Ann.*, 20 (1882) 213–225.
24. G. Miel, On calculations past and present: the Archimedean algorithm, *Amer. Math. Monthly*, 90 (1983) 17–35.
25. D. J. Newman, Rational Approximation Versus Fast Computer Methods, in Lectures on Approximation and Value Distribution, Presses de l'Université de Montreal, 1982, pp. 149–174.
26. S. Ramanujan, Modular equations and approximations to $\pi$, *Quart. J. Math*, 45 (1914) 350–72.
27. E. Salamin, Computation of $\pi$ using arithmetic-geometric mean, *Math. Comput.*, 30 (1976) 565–570.
28. B. Schoenberg, Elliptic Modular Functions, Springer, Berlin, 1976.
29. A. Schönhage and V. Strassen, Schnelle Multiplikation Grosser Zahlen, *Computing*, 7 (1971) 281–292.
30. D. Shanks, Dihedral quartic approximations and series for $\pi$, *J. Number Theory*, 14 (1982) 397–423.
31. D. Shanks and J. W. Wrench, Calculation of $\pi$ to 100,000 decimals, *Math Comput.*, 16 (1962) 76–79.
32. W. Shanks, Contributions to Mathematics Comprising Chiefly of the Rectification of the Circle to 607 Places of Decimals, G. Bell, London, 1853.
33. Y. Tamura and Y. Kanada, Calculation of $\pi$ to 4,196,393 decimals based on Gauss-Legendre algorithm, preprint (1983).
34. J. Tannery and J. Molk, Fonctions Elliptiques, vols. 1 and 2, 1893; reprint Chelsea, New York, 1972.
35. S. Wagon, Is $\pi$ normal?, *The Math Intelligencer*, 7 (1985) 65–67.
36. G. N. Watson, Some singular moduli (1), *Quart. J. Math.*, 3 (1932) 81–98.
37. ______, The final problem: an account of the mock theta functions, *J. London Math. Soc.*, 11 (1936) 55–80.
38. H. Weber, Lehrbuch der Algebra, Vol. 3, 1908; reprint Chelsea, New York, 1980.
39. E. T. Whittaker and G. N. Watson, A Course of Modern Analysis, 4th ed, Cambridge Univ. Press, London, 1927.

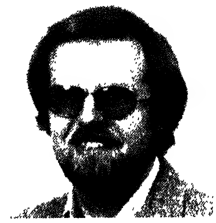
# Equations in Division Rings—A Survey

J. LAWRENCE*, *University of Waterloo*

G. E. SIMONS, *Royal Military College*



JOHN LAWRENCE was born in Ottawa, Ontario, and studied mathematics at Carleton University and McGill University. He now teaches at the University of Waterloo.



GORDON SIMONS was born in Calgary, Alberta, and studied mathematics at the University of Waterloo and the University of Toronto. He received a Ph.D. in Pure Mathematics from the University of Waterloo in 1983. His major fields of interest are ring theory, especially noncommutative ring theory, and its connections with universal algebra. He teaches both mathematics and computer science at the Royal Military College of Canada, where he is an assistant professor.

**I. Introduction.** One of the most basic, yet difficult, questions in mathematics is: can you solve this equation? Many adjectives can be inserted before the word "equation"; in this article we are interested in polynomial equations over division rings.

A great deal of mathematics has arisen from attempts to solve polynomial equations, or to prove them unsolvable, ranging from the formulas for roots of cubic and quartic polynomials of the Italian algebraists of the 16th century to Galois' theory of solvability, which led to modern field theory and group theory.

For polynomial equations with coefficients from a field, results on the nature and existence of solutions are very well known. It was known to the Greeks that there were equations with rational coefficients, such as $x^2 - 2 = 0$, with no rational roots, while Gauss published his first proof of the Fundamental Theorem of Algebra (every polynomial with complex coefficients has a root in the complex numbers) in 1799.

The results surveyed in this paper are all, in some sense, motivated by one question: do theorems on polynomial equations over fields still hold for polynomial equations over division rings (noncommutative fields)?

We will see that the results about equations over division rings are often surprisingly different from the familiar results for equations over fields and that some of the most well-known results about equations over fields fail quite dramatically for equations over division rings. For other results, there are division ring

*Research supported by a grant from NSERC

analogues which reduce to the field case in the presence of commutativity. In general, results about the solution of equations over division rings are much less complete than the results for equations over fields.

Examples of division rings are less familiar than examples of fields. The real quaternions, denoted $\mathbb{H}$ and discovered by Hamilton in 1843, are certainly the best known example of a (noncommutative) division ring. $\mathbb{H}$ is a 4-dimensional real vector space with basis $\{1, i, j, k\}$ and multiplication determined by the identities $i^2 = j^2 = k^2 = -1$, $k = ij = -ji$. This construction can be generalized by replacing $\mathbb{R}$ by other fields.

Some other examples of division rings are more difficult to describe explicitly, such as quotient rings of skew polynomial rings, crossed products such as cyclic division algebras and generic (or universal) division rings. Details on these and other examples of division rings can be found in [Cohn, 1977] or [Dauns, 1982].

Throughout this article $D$ will denote a division ring, $Z(D)$ its centre (a field in its own right) and $F$ a field. Other terms used for division rings include skew field, sfield and, most confusingly, field. Those adopting this last term usually (but not always) refer to commutative fields when they mean a field (in our sense).

A division ring $D$ is a vector space over its centre $Z(D)$. We denote $[D : Z(D)]$ by dim $D$, the dimension of $D$ as a vector space over $Z(D)$ and write dim $D = \infty$ if this dimension is infinite. In the study of division rings, the finite dimensional and infinite-dimensional cases often require quite different techniques.

At times, we may want to consider a class of division rings which all have the same field contained in their centres. A ring $R$ is an $F$-algebra if $R$ has a copy of the field $F \subseteq Z(R)$. Note that $Z(R)$ may be larger than $F$.

Before beginning our survey of polynomial equations over a division ring, we have to consider one more question: exactly what do we mean by a polynomial equation over a division ring?

It seems clear that a polynomial equation over the division ring $D$ is an equation of the form $f(x_1, x_2, \ldots, x_n) = 0$ where $f$ is a polynomial over $D$, but there are a number of different interpretations of the notion of a polynomial over $D$.

In the case of an equation in one variable, $f(x) = 0$, one possibility is to take $f$ to be of the form $a_n x^n + a_{n-1} x^{n-1} + \cdots + a_n x + a_0$ with the coefficients $a_i$ all in $D$. The set of all such expressions will be denoted $D_L[x]$, following the notation of [Gordon and Motzkin, 1965]. $D_L[x]$ can be made into a ring by defining addition in the usual way and defining multiplication as though $x$ were central. If we do this, the evaluation map $D_L[x] \to D$, $x \mapsto a$, $a \in D$ is not a ring homomorphism unless $a \in Z(D)$. Thus we can have a factorization $f(x) = g(x)h(x)$ in $D_L[x]$ without necessarily having $f(a) = g(a)h(a)$ for all $a \in D$.

A more general notion of polynomial is to take $f$ to be an element of the coproduct $D *_F F[x]$, where $F \subseteq Z(D)$ (often $F = Z(D)$). An element of $D *_F F[x]$ is a finite sum of monomials of the form $a_1 x a_2 x a_3 x \cdots a_n x a_{n+1}$, where the $a_i \in D$. Let $\mathscr{B}$ be a basis of $D$ over $F$ with $1 \in \mathscr{B}$. Each element of $D *_F F[x]$ can be expressed uniquely as a linear combination (over $F$) of monomials of the form $a_1 x a_2 x \cdots x a_{n+1}$, where $a_i \in \mathscr{B}$. This gives a "normal form" for elements of $D *_F F[x]$. The degree of a monomial is the number of $x$'s that appear in it and the degree of a polynomial is the maximum of the degrees of the monomials in it. This is independent of the basis of $D$ over $F$. In $D *_F F[x]$, elements of $F$ commute with $x$ while elements of $D \setminus F$ do not commute with $x$. Thus the expression $ax - xa$ is

equal to 0 exactly when $a \in F$, so that if $a \in D \setminus F$ then $ax - xa$ is a polynomial (element of $D *_F F[x]$) composed of two monomials and has degree one.

For polynomial equations in more than one variable, $f(x_1, \ldots, x_n) = 0$, we take $f \in D *_F F\langle x_1, \ldots, x_n \rangle$. $F\langle x_1, \ldots, x_n \rangle$ denotes the free $F$-algebra on $\{x_1, \ldots, x_n\}$, so the indeterminates commute with $F$ but not with each other. An element of $D *_F F\langle x_1, \ldots, x_n \rangle$ is then a sum of monomials of the form $a_1 x_{i_1} a_2 x_{i_2} a_3 \ldots a_k x_{i_k} a_{k+1}$. The degree of such a monomial is $k$ and the degree of a polynomial is the maximum of the degrees of its (nonzero) monomials.

A solution of the polynomial equation $f(x_1, \ldots, x_n) = 0$ is simply an $n$-tuple of elements of $D$, $(d_1, \ldots, d_n)$ such that $f(d_1, d_2, \ldots, d_n) = 0$. The solution set of the equation is the set of all solutions of the equation. Two different equations may have the same solution set since one equation may be transformed into another (by various means).

At times we may wish to compare the polynomial equations that are satisfied by two different division rings, so the kinds of polynomials considered must make sense in both rings. If both rings are $F$-algebras for some field $F$ then we could consider elements of the free algebra $F\langle x_1, \ldots, x_n \rangle$ as the polynomials to be used.

If we wish to compare equations satisfied by any two division rings, then elements of the free ring $\mathbb{Z}\langle x_1, \ldots, x_n \rangle$ can be used. In this case, a $\mathbb{Z}$-polynomial is often called a (ring) term and a polynomial equation is called a term equation. This terminology is consistent with the usage in universal algebra.

**II. Single polynomial equations.** In the study of polynomial equations over a division ring, it seems very natural to attempt to produce analogues of the classical results for polynomials over a field. What are the division ring versions of theorems such as: every linear equation over a field has a unique solution in the field, every polynomial of degree $n$ over a field has $n$ roots in some extension of the base field, a root of a polynomial gives a factorization of the polynomial? We shall see that the division ring results are often vastly different from the results for fields. In many cases, there are no satisfactory analogues and known results are far from complete.

The first case we examine is that of linear equations. Over a field the situation is trivial: $ax = b$ has a unique solution in the field if and only if $a \neq 0$.

For an example of how different the situation can be in a division ring, consider the equation $ix - xi = 1$ over $\mathbb{H}$, the real quaternions. Not only does this equation have no solution in $\mathbb{H}$, it cannot have a solution in any ring containing $\mathbb{H}$, since if there were such an $x$, $i^2x = -x$ but $ix = 1 + xi$ gives $i^2x = i(1 + xi) = i + (ix)i = i + (1 + xi)i = 2i - x$, and so $2i = 0$, a contradiction.

Over a division ring, the general linear equation has the form $\sum_{i=1}^{n} a_i x b_i = c$, where $n$ is a positive integer. Most efforts on solving linear equations over division rings have concentrated on the solution of the equation $ax - xb = c$. Clearly if either $a$ or $b$ is central and $a \neq b$ then there is a unique solution in the division ring, while if $a = b$ and $a$ is central, then there is no solution unless $c = 0$, in which case the equation is just $0x = 0$. Thus one condition to consider is the following: for a ring $R$

$$\text{the equation } ax - xb = c, \quad a \neq b,\ a, b, c \in R \text{ has a solution in } R. \qquad (*)$$

It is easy to see that any ring satisfying (*) must be a division ring. Meisters [1961] has shown that if we also require the solution to $ax - xb = c$ to be unique, then the

ring must be a field. He also proved that any division ring finite dimensional over its centre that satisfies (*) must be a field. Lazerson [1961] states (without proof) that if $D$ is a division ring satisfying (*), then every non-central element of $D$ must be transcendental over the centre of $D$. Thus a condition like (*), while a natural analogue of the situation over a field, seems to be too restrictive in the case of division rings.

Instead let us ask for conditions on elements $a, b, c$ of a division ring $D$ so that the equation $ax - xb = c$ has a solution. Since this is a linear equation, its solution will be unique if and only if the equation $ax - xb = 0$ has $x = 0$ as the only solution. Thus if $a$ and $b$ are not conjugate, any solution to $ax - xb = c$ will be unique. If $a$ and $b$ are conjugate then there will not be a unique solution. In this case the equation can be transformed into $ay - ya = d$, since $b = tat^{-1}$ for some nonzero $t$, so $ax - xb = c = ax - xtat^{-1}$ gives $a(xt) - (xt)a = ct$.

If the division ring $D$ is algebraic and separable over its centre, Johnson [1944] proved:

THEOREM. *If $a$ and $b$ are not conjugate, then $ax - xb = c$ has a unique solution in $D$, where $a, b, c \in D$.*

He also found necessary and sufficient conditions for a solution when $a$ and $b$ are conjugate.

A special case of this result is the following.

THEOREM. *If $D$ is a quaternion algebra over a formally real field $F$ (i.e., $-1$ is not a sum of squares in $F$) and $a \in D \setminus F$ then $ax - xa = b$ has a solution in $D$ if and only if $ba = \bar{a}b$.*

Another way to approach the solution of $ax - xa = b$ is to consider the derivation map $x \mapsto ax - xa$. Using this approach Jacobson [1944] found necessary and sufficient conditions for the equation $ax - xa = b$ to have a solution where $a$ is an algebraic element in an arbitrary algebra. Lazerson [1961] gives a construction for a division ring that has an element $t$ whose associated derivation $x \mapsto xt - tx$ is onto, so that the equation $xt - tx = b$ has a solution for all elements $b$.

We have seen that $ax - xb = c$ does not always have a solution in a given division ring $D$. Can $D$ be embedded in another division ring in which $ax - xb = c$ is solvable?

P. M. Cohn in [Cohn, 1958, 1959] considered this question and was able to prove the following result in this direction.

THEOREM. *Let $A$ be an algebra over a field $F$, without zero-divisors. Then $A$ can be embedded in an algebra $A^*$ over $F$, without zero-divisors, such that for any two elements $a, b \in A^*$, neither algebraic over $F$, and any $c \in A^*$ there is an $x \in A^*$ with $ax - xb = c$.*

While the algebra $A$ can be chosen to be a division ring, with $F$ the centre of $A$, there is no guarantee that the algebra $A^*$ will be a division ring, although clearly all algebraic elements of $A^*$ are invertible.

For a linear equation over a division ring, the most complete results have been obtained by Cohn [1973b]. In that paper, he proved the following theorem.

THEOREM. *Let $D$ be a division ring and $F$ a fixed subfield of $Z(D)$. Consider the equation $ax - xb = c$, where $a, b, c \in D$. Then*

(i) *if $a$ and $b$ are both transcendental over $F$ then there is a division ring $E$, containing $D$ with $F \subseteq Z(E)$, which has infinitely many solutions of the equation,*

(ii) *if one of $a$ or $b$ is algebraic over $F$ while the other is either transcendental over $F$ or algebraic over $F$ but with a different minimal equation, then there is a unique solution in $D$ of the equation,*

(iii) *if $a$ and $b$ have the same minimal equation over $Z(D)$ then there is a solution of $ax - xb = c$ in an extension of $D$ if and only if either $c = 0$ or $(t - cbc^{-1}) \times (t - a)$ divides the minimal polynomial of $a$ over $Z(D)$.*

Conditions (ii) and (iii) imply the earlier results of Johnson for elements of an algebraic division ring.

As an example of the use of this theorem consider the "metro-equation" $ax - xa = 1$ over a division ring $D$ with centre $F$. If $a$ is transcendental over $F$ then there is a division ring extension of $D$ which contains a solution. If $a$ is algebraic and separable over $F$ there is no solution in $D$ or any extension of $D$, while if $a$ is algebraic but not separable over $F$ there is a solution in $D$.

For the metro-equation, there is an interesting short proof that there is no solution when $a$ is algebraic and separable over $F$. Suppose that $a \in D \setminus F$ is algebraic over $F$ with minimal polynomial $f$. It follows by induction that $a^n x - xa^n = na^{n-1}$ for $n \geqslant 1$ and hence that $f(a)x - xf(a) = f'(a) = 0$ as $f(a) = 0$. Thus $f' = 0$ by the minimality of $f$. If $a$ is separable over $F$, then $a \in F$, a contradiction.

In the case of polynomials of degree greater than one over a division ring there are even fewer results known than in the linear case. We would like to find results similar to those known for polynomial equations over a field, namely that a polynomial $p$ over a field has a root in some extension field, $p(a) = 0$ implies that $p(x) = q(x)(x - a)$ for some polynomial $q$ (the Factor Theorem) and hence that a polynomial of degree $n$ has $n$ roots in a suitable extension field.

To illustrate the problems which can arise when working over division rings, consider the equation $x^n - a = 0$ where $n \geq 2$. If we work in the ordinary real quaternions $\mathbb{H}$, then if $a \notin \mathbb{R}$ it has exactly $n$ $n$th roots in $\mathbb{H}$. If $a$ is real then it has infinitely many $n$th roots, unless $n = 2$ and $a > 0$, in which case there are exactly two square roots. For example $x^2 + 1 = 0$ has solutions $(ai + bj + ck) / \sqrt{a^2 + b^2 + c^2}$, where $a, b, c \in \mathbb{R}$ and $a^2 + b^2 + c^2 \neq 0$, so $-1$ has uncountably many square roots in $\mathbb{H}$. Similarly, 1 has uncountably many cube roots in $\mathbb{H}$, since $x^3 - 1 = 0$ has solutions $1, -1/2 + (\sqrt{3}/2)\alpha, -1/2 - (\sqrt{3}/2)\alpha$, where $\alpha \in \mathbb{H}$ and $\alpha^2 = -1$. These results are proved in [Niven, 1942] and [Brand, 1942]. Brand's note shows that the calculations are essentially the same as finding $n$th roots over the complex numbers.

Fein and Schacher [1971] consider the problem of characterizing the rational numbers which have $n$th roots in some finite-dimensional division ring extension of $\mathbb{Q}$. They show, for example, that if $n$ is not congruent to 0 mod 4, then there is a division ring $D$, finite dimensional over $\mathbb{Q}$, containing an $n$th root of $a$, while $-25$ has no 4th roots in any such $D$.

For equations of the form $x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0$ with all coefficients in a division ring $D$, Niven [1941] showed that if $D$ is a quaternion algebra

over a real closed field then such an equation has a root in $D$. In particular, any such equation over the real quaternions $\mathbb{H}$ has a root in $\mathbb{H}$.

This property is unique to quaternion division algebras (among finite-dimensional division rings) because of a result of Baer, given in Niven's paper, that the only finite-dimensional division algebras $D$ for which $x^n + a_{n-1}x^{n-1} + \cdots + a_0 = 0$, $a_i$'s $\in D$, always has a solution in $D$ are quaternion algebras over real closed fields (a field $F$ is real closed if $-1$ is not a sum of squares in $F$, but is a sum of squares in every proper algebraic extension of $F$). Thus in a sense the real quaternion algebra is "algebraically closed."

More recently, Cohn [1975a] has shown that every equation $x^n + a_{n-1}x^{n-1} + \cdots + a_0 = 0$ with coefficients in a division ring $D$ has a solution in some division ring extension of $D$.

For more general polynomial equations $f(x) = 0$ with $f \in D *_F F[x]$, the results are more fragmented. Eilenberg and Niven [1944] showed that if $f \in \mathbb{H} *_{\mathbb{R}} \mathbb{R}[x]$ and $f$ has only one monomial of maximum degree, then $f(x) = 0$ has a solution in $\mathbb{H}$. Their proof uses algebraic topology, a homotopy argument and the existence of $n$th roots in $\mathbb{H}$ to deduce the existence of a root of $f(x) = 0$.

Cohn has pointed out that a result of R. Wood [1985] leads to the following more general result. If $D$ is a quaternion algebra over a real closed field $F$ and $f \in D *_F F[x]$ has only one monomial of maximum degree (greater than zero), then $f(x) = 0$ has a solution in $D$.

Cohn [1975a] has proved some results for equations of certain forms or equations over particular division rings. For example, the equation $xax + bx + xc + d = 0$, $a \neq 0$, always has a solution in some extension.

If every polynomial equation (of nonzero degree) over $D$ has a solution in some extension of $D$, then there are some restrictions on the structure of $D$.

THEOREM [Cohn, 1977]. *Let $D$ be a division ring with centre $F$. If every nonconstant polynomial over $D$ has a root in some extension of $D$, then $F$ is algebraically closed in $D$.*

The metro-equation shows that $F$ is separably closed in $D$. A different argument is used to show that there are no elements purely inseparable over $F$ (see [Cohn, 1977], p. 201).

Cohn has connected the solving of polynomial equations in division ring extensions to the solution of matrix equations and the existence of singular eigenvalues. This has led to the following:

CONJECTURE [Cohn, 1973a]. *Let $D$ be an $F$-division algebra with $F$ algebraically closed in $D$. Then each nonconstant polynomial has a root in some division ring extension of $D$.*

What else can be said about the number of solutions of a polynomial equation over a division ring? For equations $f(x) = 0$ with $f \in D_L[x]$ there are some results similar to those for equations over a field. In an early paper on equations over division rings, Richardson [1927] showed that the roots of a polynomial of degree $n$ lie in at most $n$ conjugacy classes. Gordon and Motzkin [1965] gave a second proof of Richardson's result and showed that if one conjugacy class has two distinct roots of the equation, then there were infinitely many roots in that conjugacy class. Thus a polynomial of degree $n$ has either at most $n$ roots or infinitely many roots.

Earlier, Herstein [1956] had shown that the number of conjugates of an element of a division ring is either one (i.e., the element is central) or is infinite. As a consequence, any polynomial in $Z(D)[x]$ has either at most $n$ roots or infinitely many roots. This result is not as general as that of Gordon and Motzkin since the polynomials considered are more restricted in form.

While the number of solutions of a polynomial equation in one variable may be infinite there are certainly some limitations on the size of the solution set. Amitsur [1965] has proved that if $f(x_1, \ldots, x_n)$ is a nonconstant polynomial in $n$-variables over a division ring $D$ infinite dimensional over its centre, then not every $n$-tuple of elements of $D$ is in the solution set of $f(x_1, \ldots, x_n) = 0$. This does not of course hold for finite-dimensional division rings; for example if $D$ is commutative, then each pair of elements is a solution of $xy - yx = 0$.

It should be mentioned that for polynomials $f \in D_L[x]$ there is a Factor Theorem, which states that $f(a) = 0$ if and only if $f(x) = g(x)(x - a)$ for some $g \in D_L[x]$, but this does not imply that such a polynomial of degree $n$ has at most $n$ roots since the factorization is in $D_L[x]$ and the evaluation map $D_L[x] \to D$, $x \mapsto a$ is not generally a homomorphism. For polynomials in $D *_F F[x]$ there is no result similar to the Factor Theorem known.

Given $n$ elements $a_1, \ldots, a_n$ of a division ring $D$ there clearly is a polynomial in $D *_F F[x]$ which has the elements as roots, namely $(x - a_1) \cdots (x - a_n)$. Beck [1979] showed that if $D$ is a generalized quaternion algebra and $a_1, \ldots, a_n$ are $n$ non-conjugate elements in $D \setminus Z(D)$ then there is a unique polynomial in $D_L[x]$ of degree $n$ whose roots are exactly $a_1, \ldots, a_n$. He also obtained a similar result when roots are allowed to have multiplicities greater than one. These results were also obtained by Bray and Whaples [1983].

Gordon and Motzkin also considered the problem of constructing polynomials in $D *_F F[x]$ having specified numbers of roots. They proved that if $[D : F] = d$ is finite, $n$ an integer greater than 0 and $h$ an integer with $0 \leq h \leq n^d$, then there is a polynomial in $D *_F F[x]$ of degree $n$ which has exactly $h$ zeroes. This differs sharply from the results known when the polynomials considered are in $D_L[x]$.

**III. Systems of equations.** We now turn to some questions about systems of equations in division rings. Two systems of equations in the same set of variables are *equivalent* over a given division ring if they have the same solution set in the division ring.

We first examine whether an infinite system of equations is equivalent to some finite subsystem. In one case, we have the following:

THEOREM A. *Over a finite-dimensional division ring, an infinite system of polynomial equations in a finite number of variables is equivalent to a finite subsystem.*

In this theorem, the term "polynomial equations" can be taken to mean either elements of $\mathbb{Z}\langle x_1, \ldots, x_n \rangle$ (term equations) or elements of $D *_F F\langle x_1, \ldots, x_n \rangle$, where $F = Z(D)$.

One consequence of this theorem is that for finite-dimensional division rings, an infinite conjunction of equations can be replaced by a single statement which is the conjunction of the equivalent finite subsystem of equations.

The proof of this theorem uses the fact that a finitely generated generic matrix ring, with coefficients in the integers, satisfies the ascending chain condition on

semiprime ideals. [Rowen, 1980] contains a proof of this fact (see Theorem 4.5.7) and related definitions.

Theorem A is not true for division rings in general. As an example, consider the division ring and infinite system of equations (in four variables) constructed as follows. Begin with the two-generated solvable torsion-free group $G$ constructed in [Baumslag and Solitar, 1962]. This group is isomorphic to a proper homomorphic image of itself. Now form the integral group ring $\mathbb{Z}G$, which is an Ore domain and embed it in a division ring $D$ [Lewin, 1972]. This is the division ring we use.

To construct the system of equations, we find a sequence of epimorphisms $\phi_1, \phi_2, \dots$ from the two-generated free group into $G$ with the property that for each $i \in \mathbb{N}$, $\ker \phi_i \subsetneqq \ker \phi_{i+1}$. Choose a group word $w_{i+1}(u, v) \in \ker \phi_{i+1} - \ker \phi_i$ and replace all occurrences of $u^{-1}$ and $v^{-1}$ by $x$ and $y$ respectively to obtain a semigroup word $w'_{i+1}(u, v, x, y)$. Then in $D$ the system of (ring) term equations $\{ w'_i(u, v, x, y) - 1 = 0;\ i \in \mathbb{N}\}$ is not equivalent to any finite subsystem.

We do not know how pathological this example is. As a result we ask the following.

*Question B.* Is there a division ring, infinite dimensional over its centre, for which Theorem A holds on term equations?

Dr. G. Revesz has informed us (oral communication) that a result of his shows that over the free field (the universal field of fractions of the free algebra) a system of term equations in two variables is equivalent to a finite subsystem.

Rather than consider systems of equations over a single division ring, we now turn to comparing systems of term equations satisfied by a pair of division rings. In particular, if we have two division rings $D$ and $D'$ such that every term equation with a solution in $D$ has a solution in $D'$, then what relationship is there between $D$ and $D'$?

To begin, consider the commutative case. If we have two fields $F$ and $F'$, both algebraic over the rationals, then every term equation which has a solution in $F$ has a solution in $F'$ if and only if $F$ is isomorphic to a subfield of $F'$. On the other hand, if two fields both contain a copy of the algebraic closure of the rationals, then they satisfy the same term equations.

In the division ring case, the analogous result is the following.

THEOREM [Lawrence]. *If $D$ and $D'$ are division rings, then every finite system of term equations which has a solution in $D$ has a solution in $D'$ if and only if $D$ can be embedded in an ultrapower of $D'$.*

A general reference for basic properties of ultraproducts is [Bell and Slomson, 1969]. In essence, ultraproducts preserve first-order logical properties. In particular, dim $D$ is preserved under ultrapowers, so by a theorem of Bergman and Small we obtain:

COROLLARY [Bergman and Small, 1975]. *If $D$ and $D'$ are division rings and every finite system of term equations which has a solution in $D$ has a solution in $D'$, then $\dim D$ divides $\dim D'$ (as integers).*

For example, the equation $(xy - yx)z = 1$ has a solution if and only if the division ring is not commutative, that is, if and only if, dim $D \neq 1$. It is rather easy to find an example of a term equation which has a solution in every division ring of

dimension 9 but has no solution in any division ring of dimension 4. Using an example of Bergman [1976] of a rational identity satisfied by division rings of dimension 9 but not satisfied by division rings of dimension 4, one can construct a system of (term) equations (in about 25 variables) which has a solution in every division ring of dimension 4 but has no solution in any division ring of dimension 9.

Conversely, if two division rings are similar algebraically, we can ask if they satisfy the same equations. One particular instance of such a question is:

*Question.* If two division rings are both of dimension four over their centres and both centres contain the algebraic closure of the rationals, do the division rings satisfy the same term equations?

Next we look at some recent applications of systems of equations over division rings to some problems in formal language theory.

Let $\mathbb{M}$ be a finitely generated free monoid with generating set $\{u_1, \dots, u_k\}$. A subset $\mathscr{L} \subseteq \mathbb{M}$ is called a *language* in the *alphabet* $u_1, \dots, u_k$. In the early 1970's Ehrenfeucht asked the following: for each language $\mathscr{L}$ does there exist a finite subset $S$ of $\mathscr{L}$ such that any two endomorphisms of the monoid $\mathbb{M}$ which agree on $S$ must agree on $\mathscr{L}$? Such a subset, if it exists, is called a *test set* for $\mathscr{L}$.

Culik and Karhumaki [Karhumaki, 1984] showed that the existence of a test set was equivalent to showing that each system of monoid equations (in a finite set of variables) is equivalent over the free monoid to a finite subsystem. This was proved recently by Albert and Lawrence using group theory [Albert and Lawrence, 1985].

Another proof uses Theorem A. First note that the free monoid on a countable set of generators can be embedded into a finite-dimensional division ring [Lichtman, 1978]. Now as monoid equations are also ring equations, Theorem A gives us the desired result.

The question of Ehrenfeucht can be generalized by allowing a finite number of substitutions rather than a single substitution (i.e., an endomorphism). For example, let $\mathbb{M}$ be the free monoid generated by $\{u, v\}$, let $\mathscr{L} = \{u^n v^n; n = 1, 2, \dots\}$, let $\phi$ map $u$ to the set $\{1, u\}$ and $v$ to the set $\{v, uv\}$ and let $\psi$ map $u$ to the set $\{u, v\}$ and $v$ to $\{v^2\}$. Then $uv$ is mapped to the multiset $\{v, uv, uv, u^2v\}$ under $\phi$ and to the multiset $\{uv^2, v\}$ under $\psi$, so that the substitutions $\phi$ and $\psi$ do not agree on $\{uv\}$. For each language $\mathscr{L}$ is there a finite test set $S$ such that two finite substitutions agreeing on $S$ (i.e., giving the same multiset) will agree on $\mathscr{L}$?

This question has an affirmative answer if every system of monoid equations is equivalent to a finite subsystem over $\mathbb{Z}\langle X\rangle^*$, the multiplicative monoid of the free ring (with zero removed). In particular, an affirmative answer to Question B would give an affirmative answer to the question on finite substitutions.

**IV. Algebraically closed division rings.** The notion of an algebraically closed field can be generalized to division rings in several ways. We will use a natural generalization due to A. Robinson [Robinson, 1971]. This defines a division ring $D$ to be *algebraically closed* if every finite system of polynomial equations over $D$ which has a solution in an extension of $D$ has a solution in $D$. We will also say that a division ring $D$ of dimension at most $n^2$ over its centre is *algebraically closed of degree n* if each finite system of polynomial equations which has a solution in an extension division ring of dimension at most $n^2$ over its (own) centre has a solution in $D$.

Thus an algebraically closed field is the same as a division ring which is algebraically closed of degree 1, and an algebraically closed division ring is the same as a division ring which is algebraically closed of degree $\infty$.

A division ring $D$ is said to be *existentially closed* if an existential sentence (a sentence of the form $\exists x_1 \exists x_2, \ldots, \exists x_n Q(x_1, \ldots, x_n)$, where $Q$ has no quantifiers) which holds in a division ring extension of $D$ must hold in $D$. It is rather easy to show that in a division ring an existential sentence (in the language of rings) is equivalent to an existential sentence in which the matrix part is equivalent to a conjunction of equations, so a division ring is existentially closed if and only if it is algebraically closed.

The remainder of this section looks at some of the properties of existentially closed division rings of degree $\infty$ and existentially closed division rings of finite degree (greater than one).

Existentially closed division rings of degree $\infty$ have been investigated in [Cohn, 1975], [Macintyre, 1974] and [Wheeler, 1972, 74]. Wheeler [1972] has found an existential sentence $P(x_1, \ldots, x_n)$ which expresses the fact that the $x_i$'s commute with each other and are algebraically independent. Thus an existentially closed division ring contains an algebraically closed subfield of infinite transcendence degree, so that, in this sense, existentially closed division rings are large. However, the centre of such a division ring is as small as possible, that is, either the prime subfield (regarding the division ring simply as a ring) or the central field $F$ (regarding the division ring as an $F$-algebra). Just as for algebraically closed groups, an existentially closed division ring cannot be finitely generated or finitely related, and the class of existentially closed division rings is not elementary. Some results on existentially closed division rings have been used by Wheeler to formulate a version of Hilbert's Nullstellensatz [Wheeler, 1972], [Cohn, 1974].

Unlike the commutative case there is really no notion of algebraic closure for division rings. For example, there is no existentially closed division ring containing the rationals such that every existentially closed division ring of characteristic 0 contains a copy of this division ring, hence there is no minimal existentially closed division ring of characteristic 0. In fact, a countable existentially closed division ring contains a proper subring isomorphic to itself, while the intersection of all its algebraically closed subfields is again as small as possible. Also, a countable existentially closed division ring has a continuum of automorphisms, so it must have outer automorphisms.

The situation does not improve if we restrict our attention to existentially closed division rings of degree $n$, where $n > 1$, as we can see from examples in the case $n = 2$.

We have seen that in some sense the real quaternion algebra is algebraically closed, but it turns out not to be existentially closed of degree 2. This follows from a theorem of Pfister [1971] which shows that there is a field extension of the real numbers in which $-1$ is a sum of four squares but not a sum of two squares. Then the system of equations

$$\begin{aligned} &w^2 + x^2 + y^2 + z^2 = -1 \\ &\qquad iw - wi = jw - wj = ix - xi = jx - xj = iy - yi = jy - yj \\ &\qquad\qquad = iz - zi = jz - zj = 0 \end{aligned}$$

has no solution in the real quaternions but it does have a solution in a division ring extension of dimension 4.

It also follows from this result that the class of (existentially closed) division rings of dimension 4 does not have the amalgamation property. For example there is no (existentially closed) division ring of dimension 4 that contains isomorphic copies of ordinary quaternion algebras over $\mathbb{Q}(\sqrt{-7})$ and $\mathbb{Q}(\sqrt{-15})$.

In spite of these examples, one may hope for some reasonable properties of existentially closed division rings of finite degree. With this hope in mind, we ask:

*Question.* If two division rings $D$ and $D'$ can both be embedded in a division ring of dimension $n^2$, are existentially closed of degree $n$ and their centres have the same transcendence degree (over the prime subfield), is $D$ then isomorphic to $D'$?

Finally we should point out that several authors have used other definitions of algebraically closed division rings which differ from Robinson's definition. Makar-Limanov [1985] looks at algebraic closure from the point of view of a single polynomial equation in one variable. He shows that there are noncommutative division rings of characteristic zero in which each nontrivial polynomial has a root. Dauns [1982] and Smith [1977] use yet another definition of algebraically closed involving factorizations of elements of $D_L[x]$.

**Acknowledgement.** The authors would like to thank P. M. Cohn for his comments and corrections.

REFERENCES

1. M. Albert and J. Lawrence, A proof of Ehrenfeucht's Conjecture, *Theo. Computer Sci.*, 41 (1985) 121–23.
2. S. A. Amitsur, Generalized polynomial identities and pivotal monomials, *Trans. Amer. Math. Soc.*, 114 (1965) 210–226.
3. G. Baumslag and D. Solitar, Some two generator one relator non-Hopfian groups, *Bull. Amer. Math. Soc.*, 68 (1962) 199–201.
4. B. Beck, Sur les équations polynomiales dans les quaternions, *Enseign Math.*, 25 (1979) 193–201.
5. J. L. Bell and A. B. Slomson, Models and Ultraproducts, North-Holland Publ. Co., 1969.
6. G. Bergman, Rational relations and rational identities in division rings I, II, *J. Algebra*, 43 (1976) 252–297.
7. G. Bergman and L. Small, PI-degrees and prime ideals, *J. Alg.*, 33 (1975) 435–462.
8. M. Boffa and P.v. Praag, Sur les corps génériques, *C. R. Acad. Sci. Paris Sér. A*, 274 (1972) 1325–1327.
9. M. Boffa, A note on existentially closed complete division rings, Springer Lecture Note 498 (1975), 56–59.
10. L. Brand, The roots of a quaternion, *Amer. Math. Monthly*, 49 (1942) 519–520.
11. U. Bray and G. Whaples, Polynomials with coefficients from a division ring, *Canad. J. Math.*, 35 (1983) 509–515.
12. S. Burris and H. P. Sankappanavar, A Course in Universal Algebra, Springer-Verlag, 1981.
13. L. Cerliénco and M. Mureddey, A note on polynomial equations in quaternions, *Rend. Sem. Fac. Sci. Univ. Cagliari*, 51 (1981) 95–99.
14. P. M. Cohn, On a class of simple rings, *Mathematika*, 5 (1958) 103–117.
15. ______, Simple rings without zero-divisors and Lie division rings, *Mathematika*, 6 (1959) 14–18.
16. ______, Factorization in general rings and strictly cyclic modules, *J. Reine Angewandte Math.*, 239/240 (1970) 185–200.
17. ______, Universal skew fields of fractions, *Symposia Mathematica*, 8 (1972) 135–148.
18. ______, Skew Field Constructions, Carleton Lecture Notes (1973).
19. ______, The range of derivations on a skew field and the equation $ax - xb = c$, *J. Indian Math. Soc.*, 37 (1973) 61–69.

20. ______, Progress in free associative algebras, *Israel J. Math.*, 19 (1974) 109–151.
21. ______, Equations dans les corps gauches, *Bull. Soc. Math. Belg.*, 27 (1975) 29–39.
22. ______, Presentations of skew fields, I. Existentially closed skew fields and the Nullstellensatz, *Math. Proc. Camb. Phil. Soc.*, 77 (1975) 7–19.
23. ______, Skew Field Constructions, London Math. Soc. Lecture Note Series 27, 1977.
24. ______, Universal Algebra, D. Reidel Publ. Co., 1981.
25. K. Culik II and J. Karhumaki, The Decidability of the DTOL sequence equivalence problem and related decision problems, University of Waterloo, preprint, 1985.
26. J. Dauns, A Concrete Approach to Division Rings, Helderman-Verlag, 1982.
27. P. K. Draxl, Skew fields, London Mathematical Society, Lecture Note Series 81, 1983.
28. S. Eilenberg and I. Niven, The "Fundamental Theorem of Algebra" for quaternions, *Bull. Amer. Math. Soc.*, 50 (1944) 246–248.
29. B. Fein and M. Schacher, Solutions of pure equations in rational division algebras, *J. Alg.*, 17 (1971) 83–93.
30. B. Gordon and T. S. Motzkin, On the zeros of polynomials over division rings, *Trans. Amer. Math. Soc.*, 116 (1965) 218–226; Correction ibid. 112 (1966) 547.
31. I. N. Herstein, Conjugates in division rings, *Proc. Amer. Math. Soc.*, 7 (1956) 1021–1022.
32. J. Hirschfield and W. H. Wheeler, Forcing, Arithmetic, Division rings, Lecture Notes in Mathematics, Springer-Verlag, No. 545, 1975.
33. N. Jacobson, The equation $x' = xd - dx = b$, *Bull. Amer. Math. Soc.*, 50 (1944) 902–905.
34. ______, Lectures in Abstract Algebra, vol. III, Van Nostrand, 1964.
35. R. E. Johnson, On the equation $x\alpha = \gamma x + \beta$ over an algebraic division ring, *Bull. Amer. Math. Soc.*, 50 (1944) 202–207.
36. J. Karhumaki, The Ehrenfeucht conjecture: a compactness claim for finitely generated free monoids, *Theo. Comp. Sci.*, 29 (1984) 285–308.
37. J. Lawrence, Strongly related rings and rational identities of division rings, *Periodica Mathematica Hungarica*, 18 (1987) 189–192.
38. E. E. Lazerson, Onto inner derivations in division rings, *Bull. Amer. Math. Soc.*, 67 (1961) 356–358.
39. J. Lewin, A note on zero divisors in group-rings, *Proc. Amer. Math. Soc.*, 31 (1972) 357–359.
40. A. I. Lichtman, Free subgroups of normal subgroups of the multiplicative group of skew fields, *Proc. Amer. Math. Soc.*, 71 (1978) 174–178.
41. A. Macintyre, The word problem for division rings, *J. Symbolic Logic*, 38 (1973) 428–436.
42. ______, On algebraically closed division rings, preprint, 1974 (unpublished).
43. L. Makar-Limanov, On algebras with one relation, *Uspeki Mat. Nauk.*, 30, No. 2 (182) (1975) 217.
44. ______, Algebra *i* Logika.
45. ______, Algebraically closed skew fields, *J. Algebra*, 93 (1985) 117–135.
46. G. H. Meisters, On the equation $ax - xb = c$ in division rings, *Proc. Amer. Math. Soc.* 12 (1961) 428–432.
47. I. Niven, Equations in quaternions, *Amer. Math. Monthly*, 45 (1941) 654–661.
48. ______, The roots of a quaternion, *Amer. Math. Monthly*, 49 (1942) 386–388.
49. O. Ore, Linear equations in non-commutative fields, *Ann. of Math.*, 32 (1931) 463–477.
50. A. Pfister, Quadratic forms over fields, *Proc. Symposia in Pure Mathematics*, 20 (1971) *Amer. Math. Soc.*, 150–160.
51. B. Pollack, The equation $\bar{t}at = b$ in a quaternion algebra, *Duke Math. J.*, 27 (1960) 261–271.
52. G. Revesz, Universal properties of generators of a variety: groups and skew fields.
53. A. R. Richardson, Equations over a division algebra, *Mess. Math.*, 57 (1927) 1–6.
54. R. F. Rinehart, The equation $e^x e^y = e^{x+y}$ in quaternions, *Rend. Circ. Mat. Palermo*, (2) 8 (1959) 160–162.
55. A. Robinson, On the notion of algebraic closedness for noncommutative groups and fields, *J. Symbolic Logic*, 36 (1971) 441–444.
56. H. Röhrl, On the zeros of polynomials over arbitrary finite dimensional algebras, *Manuscripta Math.*, 25 (1978) 359–390.
57. L. Rowen, Polynomial Identities in Ring Theory, Academic Press, 1980.
58. A. Schofield, Representations of Rings Over Skew Fields, London Math. Soc. Lecture Note Series 92, Cambridge University Press, 1985.
59. K. Smith, Algebraically closed noncommutative polynomial rings, *Comm. in Alg.*, 5 (1977) 331–346.

60. W. H. Wheeler, Algebraically closed division rings, forcing and the analytical hierarchy, thesis, Yale University, 1972.
61. ______, Model-companions and definability in existentially complete structures, *Isr. J. Math.*, 25 (1976) 305–330.
62. R. M. W. Wood, Quaternionic eigenvalues, *Bull. London Math. Soc.*, 17 (1985) 137–138.

---

## Letters to the Editor

Editor:

One of the examples in the paper by Arora, Goel, and Rodriguez (Special techniques for trigonometric integrals, this MONTHLY (February, 1988) 126–130) is incomplete, for convergence of the improper integral was not proved. Thus all that can be concluded is that either $2I = -\pi \ln 2$ or $2I = -\infty$.

Moreover, while it is true that Ahlfors and several other textbook authors derive the formula for $\int_0^\pi \ln(\sin\theta)\,d\theta$ as an offshoot of an intricate contour integration, essentially the same presentation as the one in the MONTHLY article (including a proof of convergence) appears in the Schaum's Outline Series book on advanced calculus by Murray Spiegel (p. 275).

Gerald Leibowitz
University of Connecticut
Storrs, Connecticut 06269

# UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

*In this department the* MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

## The Missing Boundary of the Blaschke Diagram

J. R. SANGWINE-YAGER
*Department of Mathematical Sciences, Saint Mary's College, Moraga, CA 94575*

In 1916 Blaschke [3] proposed mapping the family of compact convex sets $\mathscr{K}$, in 3-dimensional Euclidean space, $E^3$, into a compact region in the plane. The range of this map is now known as the Blaschke Diagram. Two well-known geometric inequalities determine part of the boundary of the Blaschke Diagram, but the inequalities which complete the boundary remain unknown. A thirty-year-old conjecture on the convex sets mapped to the missing boundary is still unresolved, although the images of these sets under Blaschke's map are easily calculated with a computer. Furthermore, the geometric inequality which these convex sets satisfy is unknown. This paper concludes with a conjectured inequality which may bound the Blaschke Diagram.

A set is convex if the line segment joining any two points of the set is contained within it. The valuations on the family of compact convex sets involved in Blaschke's map are volume, surface area and integral mean curvature, $\int \frac{1}{2}(\kappa_1 + \kappa_2)\, dS$, where $\kappa_1$, $\kappa_2$ are the principal curvatures. For each $K$ in $\mathscr{K}$, these valuations will be denoted by $V(K) = V$, $S(K) = S$ and $M(K) = M$, respectively. The unit ball centered at the origin is denoted by $B$.

Inequalities of Minkowski state

$$S^2 \geqslant 3VM, \tag{1}$$

$$M^2 \geqslant 4\pi S, \tag{2}$$

and

$$M^3 \geqslant 48\pi^2 V. \tag{3}$$

If $K$ is a planar convex set, its area is $(1/2)S$ and its perimeter is $(2/\pi)M$. The isoperimetric inequality for planar sets becomes

$$2M^2 \geqslant \pi^3 S, \quad \text{where } V = 0. \tag{4}$$

Blaschke proposed that each member of $\mathscr{K}$ be mapped to the point $(x, y)$, where

$$x = \frac{4\pi S}{M^2} \quad \text{and} \quad y = \frac{48\pi^2 V}{M^3}.$$

The range of this map is the Blaschke Diagram and is denoted by $\bar{\mathscr{K}}$. The

inequalities above translate to

$$x^2 \geqslant y, \tag{1$'$}$$

$$1 \geqslant x, \tag{2$'$}$$

$$1 \geqslant y, \tag{3$'$}$$

and

$$\frac{8}{\pi^2} \geqslant x, \quad \text{where} \quad y = 0. \tag{4$'$}$$

Thus $\bar{\mathscr{K}}$ is a bounded region, see FIGURE 1, but the boundary from $((8/\pi^2), 0)$ to $(1, 1)$ is not known. Surveys of the problem by Hadwiger [**6**, 72–82] and Bieri [**2**] include bibliographies; contributions by both authors will be presented below.

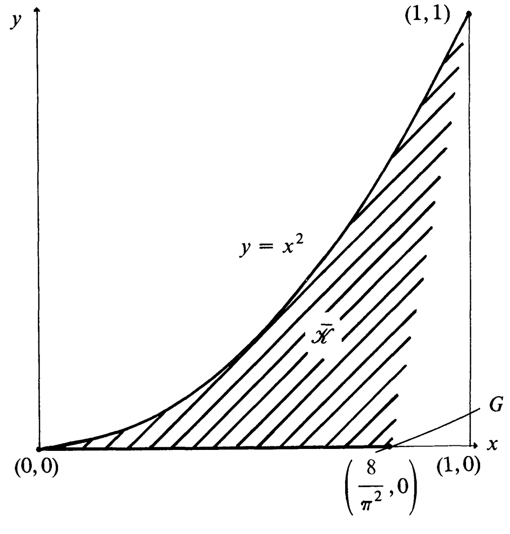



FIG. 1

Not only is a boundary of the Blaschke Diagram missing, there is only one known inequality which eliminates any part of the line $x = 1$, for $0 < y < 1$. It is due to Groemer [**4**] and states

$$V \geqslant \frac{\pi S}{24M}\left(S - \frac{2M^2}{\pi^3}\right) \tag{5}$$

or

$$y \geqslant \frac{\pi}{8}x\left(x - \frac{8}{\pi^2}\right),$$

and at

$$x = 1,\ y = 0.07439.$$

This parabola is labelled '$G$' in FIGURE 1.

Before we discuss the known extreme sets, observe that all similar sets are mapped to the same point. Examples will arise in the discussion showing that the preimage of a point may also contain nonsimilar sets. The balls are the only bodies at the point $(1,1)$. The family of extreme sets mapped to the parabolic boundary are the cap bodies. These are the convex hull of a ball and countably many points exterior to it such that the line segment joining any two of these points intersects the ball. The simplest example is a dunce cap sitting on a ball. The image of points and line segments is $(0,0)$. Planar convex sets are mapped to the x-axis, and the disks are the only sets whose image is $((8/\pi^2),0)$.

Some comments on the interior of $\mathscr{K}$ may be useful. To show that the Blaschke Diagram is connected, consider the convex array $\theta K + (1-\theta)L$, where $K$ and $L$ are elements of $\mathscr{K}, 0 \leqslant \theta \leqslant 1$, and $+$ represents Minkowski or vector addition of sets. Each element of the array is a member of $\mathscr{K}$, and it follows from the properties of $V$, $S$ and $M$, see [**6**, p. 45], that the image of the array is an algebraic curve. In particular, the image of a set $K$ is connected to $(1,1)$ by the curve determined by its outer parallel bodies (that is, bodies of the form $K + tB$, $t \geqslant 0$). If the image of $K$ lies on the line $y = 3x - 2$, then all its outer parallel bodies are mapped to that line. The curves generated by the outer parallel bodies of members of $\mathscr{K}$ mapped to the left (right) of the line have a positive (negative) second derivative. In all cases these curves have slope 3 at $(1,1)$. The curves determined by the outer parallel bodies of a line segment and a disk are labelled '$A$' and '$B$', respectively, in FIGURE 2. (For clarity, the scale on the $x$-axis in Figure 2 has been doubled.) Interestingly, the sets of constant width in $E^3$ are also mapped to a closed interval on the line $y = 3x - 2$. This follows from an identity due to Blaschke: if $K$ has constant width $b$, then $2V = bS - (2/3)\pi b^3$. The planar set of constant width having minimal area is the Reuleaux triangle, its 3-dimensional counterpart is unknown.

Two well-known conjectures have been posed about the missing boundary. Hadwiger [**5**] showed that of all sets of revolution with fixed surface area and

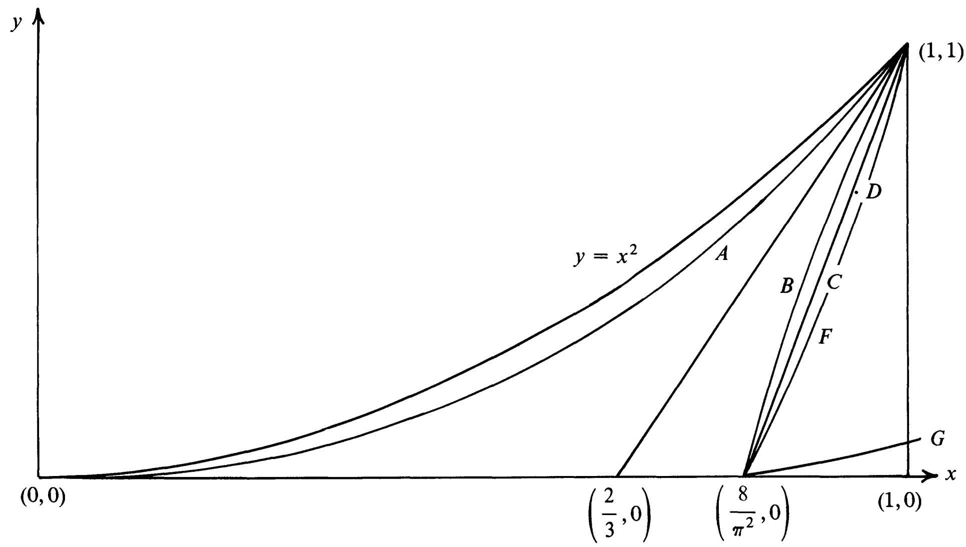



FIG. 2

mean-width, the centrally symmetric truncated spheres have the smallest volume. The image of these sets is labelled '$C$' in FIGURE 2, and Hadwiger conjectured that it was the missing boundary. A truncated sphere is a sphere with a countable number of nonintersecting segments removed. By a segment we mean the nonempty intersection of an open halfspace with a sphere which does not include the sphere's center. Bieri [**1**] showed that this conjecture was false by exhibiting a body, whose image is labelled '$D$', outside Hadwiger's boundary. This body is an asymmetric truncated sphere; three congruent segments of maximal size are removed. This led to a conjecture by Bieri which remains unresolved: the missing boundary is the image of fully truncated spheres. A sphere is fully truncated when no relatively open set on the surface of the original sphere remains. This conjecture is appealing, in part because the truncated spheres are the polar duals of the cap bodies.

While Bieri's conjecture is not resolved, much can be learned about the fully truncated spheres. Formulae may be derived for $V$, $S$ and $M$: the volume is the sum of the volume of cones, the surface area is the sum of the areas of circular faces and the total mean curvature is equal to the sum of the products of the perimeters of the faces and the angle subtended by the face at the sphere's center. The simplest way to generate a full truncation is by an osculatory packing. In such a recursive packing, a new circle is inscribed in the triangular region created by three mutually tangential circles (on a sphere or in the plane). Wilker [7] has extended the Descartes circle formula (often attributed to Soddy) to hyperbolic and spherical geometries. If three mutually tangential circles have curvatures $c_1$, $c_2$, $c_3$, and $\kappa$ is the curvature of the space, then the curvature of the inscribed circle is

$$c = c_1 + c_2 + c_3 + 2\sqrt{c_1c_2 + c_2c_3 + c_3c_1 - \kappa}\,.$$

With these formulae one may computer generate the images in the Blaschke Diagram of a multitude of fully truncated spheres. The author's efforts have been disappointing. The images lie in a band; the centrally symmetric bodies appear to be more extreme than close noncentrally symmetric ones. These bodies do not appear to be significantly more extreme than Bieri's original counterexample which is neither fully truncated nor centrally symmetric.

Finally, the author is unable to resist the temptation to contribute a conjecture. It follows from (1) and (2) that

$$\frac{M}{4\pi} \geqslant \frac{S}{M} \geqslant \frac{3V}{S}.$$

Consideration of valid geometric inequalities involving weighted arithmetic means (Bonnesen's inequality in the plane, for example) leads to the conjecture that, for all members of $\mathscr{K}$,

$$\frac{S}{M} \leqslant \frac{8}{\pi^2}\frac{M}{4\pi} + \left(1 - \frac{8}{\pi^2}\right)\frac{3V}{S}.$$

In the case of equality, this relationship translates to the parabola through $(0,0)$, $((8/\pi^2),0)$ and $(1,1)$ in the Blaschke Diagram ('$F$' in Fig. 2) and all of the images of convex sets calculated by the author are bounded by it.

REFERENCES

1. H. Bieri, Mitteilung zum Problem eines konvexen Extremalkörpers, *Arch. Math.*, **1** (1948–49) 462–463.
2. H. Bieri, Extremalkörpers rotationssymmetrischer Kugelstumpfe im gewohnlichen Raum I; II, *Elem. Math.* 33 (1978) no. 1, 7–14; no. 3, 62–68.
3. W. Blaschke, Eine Frage über konvexe Körper, *Jahresber. Deutsch. Math.-Verein.*, 25 (1916) 121–125.
4. H. Groemer, Eine neue Ungleichung für konvexe Körper, *Math. Z.*, 86 (1965) 361–364.
5. H. Hadwiger, Beweis einer Extremaleigenschaft der symmetrischen Kugelzone, *Portugaliae Math.*, 7 (1948) 74–85.
6. H. Hadwiger, Altes und Neues über konvexe Körper, Birkhäuser, Basel and Stuttgart, 1955.
7. J. B. Wilker, Inversive geometry, The Geometric Vein: the Coxeter Festschrift, Springer-Verlag, New York, 1981, pp. 379–442.

# NOTES

EDITED BY DENNIS DETURCK, DAVID J. HALLENBECK, AND ANITA E. SOLOW

## A Geometrically Inspired Proof of the Singular Value Decomposition

S. J. BLANK and NISHAN KRIKORIAN
*Department of Mathematics, Northeastern University, Boston, MA 02115*

DAVID SPRING
*Department of Mathematics, York University, Toronto, Ontario, Canada M4N 3M6*

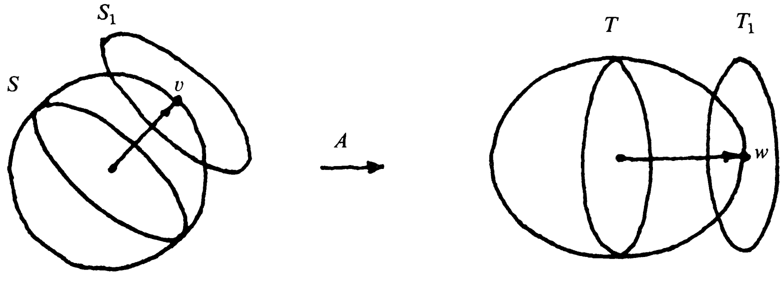
Any real $n \times m$ matrix $A$ can be factored into $A = USV'$ where $U$ and $V$ are $n \times n$ and $m \times m$ real orthogonal matrices and $S$ is a diagonal $n \times m$ matrix for which $s_{11} \geqslant s_{22} \geqslant \cdots \geqslant s_{pp} \geqslant 0$ and $p = \min(m, n)$. This is the singular value decomposition (SVD) of $A$ and the $s_{ii}$ are its singular values.

An equivalent formulation of the SVD for linear transformations can be given as follows. For any linear transformation $A\colon R^m \to R^n$ there exist orthonormal bases $v_1, v_2, \dots, v_m$ in $R^m$ and $u_1, u_2, \dots, u_n$ in $R^n$ such that $Av_i = s_{ii}u_i$ for $i = 1, \dots, p$ and $Av_i = 0$ for $i = p+1, \dots, m$. This can be viewed geometrically by saying that $A$ takes the coordinate system of $v$'s to the coordinate system of $u$'s with expansions and contractions along corresponding coordinate directions given by the $s_{ii}$'s.

The purpose of this note is to give a proof of the SVD that exposes its geometric content.

First consider the nonsingular case of $A\colon R^k \to R^k$. The image of the unit sphere $\|v\| = 1$ under $A$ is the level surface of a quadratic form. To verify this let $Av = u$, so that $1 = \|v\|^2 = \|A^{-1}u\|^2 = u'A^{-1\prime}A^{-1}u$. Moreover, since $A^{-1\prime}A^{-1}$ is positive definite, the image of the unit sphere under $A$ is an ellipsoid, and we will refer to it as such. We will never use this fact, however, since its demonstration requires a spectral resolution of the quadratic form into a positive weighted sum of squares.

Let $w$ be a radial vector of maximal length in the ellipsoid and let $v$ be its preimage under $A$. That is, $w$ is any vector such that $\|A^{-1}w\| = 1$ and $\|w\| = \max\{\|u\| : \|A^{-1}u\| = 1\}$, and $v$ is any vector such that $Av = w$. If we can show that $A$ takes the subspace $S$ perpendicular to $v$ to the subspace $T$ perpendicular to $w$, we can set $v_1 = v$, $u_1 = w/\|w\|$, and $s_{11} = \|w\|$. The proof will then continue by an obvious induction on the two subspaces.



To this end consider the plane $S_1$ that touches the sphere only at the point $v$. It is a well-known geometric property of spheres that $S_1$ must be the unique such plane and must also be perpendicular to the vector $v$. Let the plane $T_1$ be the image of $S_1$ under $A$. Since $A$ is one-one and since $S_1$ touches the sphere only at the point $v$, then $T_1$ must touch the ellipsoid only at the point $w$. In fact, $T_1$ must be the unique plane with this property. If there were another, then an application of $A^{-1}$ would produce a plane different from $S_1$ also touching the sphere only at the point $v$, an impossibility.

Furthermore, $T_1$ must be perpendicular to the vector $w$. To see this, take a large sphere centered at the origin, and shrink it until it first touches the ellipsoid. The point $w$ must occur among the points of first contact. Now consider the plane that touches the large sphere only at the point $w$. As above, this plane is unique and perpendicular to the vector $w$. Furthermore, since the large sphere completely contains the ellipsoid, this plane also touches the ellipsoid only at the point $w$ and must therefore coincide with $T_1$.

The proof for the nonsingular case then concludes with the observation that linear maps preserve parallel planes. Since $S$ is parallel to $S_1$ and $T$ is parallel to $T_1$ and since $A$ takes $S_1$ to $T_1$, it must therefore take $S$ to $T$.

Now for the general case of $A$: $R^m \to R^n$. First restrict $A$ to the orthogonal complement of its null space in $R^m$. $A$ is nonsingular on this subspace [**2**]. Use the procedure above to obtain orthonormal bases $v_1, v_2, \dots, v_k$ of this subspace in $R^m$ and $u_1, u_2, \dots, u_k$ of the range space of $A$ in $R^m$. Then extend these orthonormal sets to orthonormal bases of $R^m$ and $R^n$. This finishes the proof.

We have tried to present a proof employing only elementary geometric ideas. By recognizing "the plane touching... only at the point..." as a tangent plane and noting the differentiability of $A$, the proof can be made even shorter but more analytic.

The crucial non-geometric (but very intuitive) idea in the proof is, of course, the use of compactness to find $w$. The ellipsoid is compact since it is the continuous image of the unit sphere. The Euclidean norm, which is also continuous, therefore realizes its maximum at some point $w$ on the ellipsoid.

The proof of the SVD in Golub and Van Loan [**1**] also uses compactness, but the inductive splitting of $A$ is obtained algebraically. Both proofs can be extended to the complex case. The usual proof of the SVD examines the spectral resolution of $A'A$ [**2**]. The spectral theorem can itself be proved using compactness [**3**]. None of these proofs exhibits the compelling geometry of the one presented here.

REFERENCES

1. G. H. Golub and C. Van Loan, Matrix Computations, Johns Hopkins University Press, Baltimore, MD, 1983.
2. G. Strang, Linear Algebra and Its Applications, Academic Press, New York, NY, 1980.
3. H. Wilf, An algorithm inspired proof of the spectral theorem in $E^n$, *Amer. Math. Monthly*, 88 (1981) 49–50.

# Sum Zero (mod $n$), Size $n$ Subsets of Integers

CRAIG BAILEY and R. BRUCE RICHTER
*Department of Mathematics, U.S. Naval Academy, Annapolis, MD 21402*

In this note, a theorem of Chevalley is used to prove the following result.

THEOREM 1. *Let $n$ be any positive integer and suppose $a_1, \ldots, a_{2n-1}$ are arbitrary integers. Then there is a subset $I$ of $\{1, 2, \ldots, 2n-1\}$ such that $\sum_{i \in I} a_i \equiv 0 \pmod n$ and $|I| = n$.*

Theorem 1 is reminiscent of the well-known fact that if $a_1, \ldots, a_n$ are arbitrary integers, then there is a nonempty subset $I$ of $\{1, \ldots, n\}$ such that $\sum_{i \in I} a_i \equiv 0$ $(\operatorname{mod} n)$. However, this result is a direct consequence of the pigeonhole principle. We have been unable to prove Theorem 1 by such elementary means.

Theorem 1 is best-possible, in the sense that if there are only $2n-2$ integers, then no collection of $n$ of them need sum to 0 (mod $n$). Specifically, take $n-1$ zeroes and $n-1$ ones.

Theorem 1 is a corollary to the following conjecture [**1**].

CONJECTURE 2. *Let $a^i = (a_1^i, a_2^i, \ldots, a_k^i)$, $i = 1, 2, \ldots, k(n-1)+1$, be integral vectors. Then there is a nonempty subset $I$ of $\{1, \ldots, k(n-1)+1\}$ such that $\sum_{i \in I} \mathbf{a}^i \equiv \mathbf{0} \pmod n$, where $\mathbf{0}$ is the vector of all zeroes.*

In the event that $n$ is a prime power, then Conjecture 2 is proved in [1] using generalizations of Chevalley's Theorem. Indeed, the proof we give of Theorem 1 in the event $n$ is prime is based on [1]. Theorem 1 is a special case of Conjecture 2, with $k = 2$ and all first coordinates 1.

THEOREM 3. (*Chevalley's Theorem* [**2**], [**3**]) *Let $p$ be a prime and let $f(x_1, \ldots, x_n)$ be a polynomial of degree $< n$ such that $f(0, \ldots, 0) \equiv 0 \pmod p$. Then there is a nonzero* (mod $p$) *solution to $f(x_1, \ldots, x_n) \equiv 0 \pmod p$.*

This theorem deserves to be better known. Its proof is not difficult and much more is known. However, for our immediate purposes, it is enough.

Our proof of Theorem 1 breaks up into two parts; the first deals with the case that $n$ is a prime and the second with the case that $n$ is composite.

Let us first assume that $n$ is a prime number $p$. Let $b$ be a quadratic non-residue modulo $p$ (i.e. the equation $x^2 \equiv b \pmod p$ has no solution.)

Set

$$f(x_1, \ldots, x_{2p-1}) = \left( \sum_{i=1}^{2p-1} x_i^{p-1} \right)^2 - b \left( \sum_{i=1}^{2p-1} a_i x_i^{p-1} \right)^2.$$

Evidently, $f(0, 0, \ldots, 0) \equiv 0 \pmod p$, so Chevalley's Theorem yields a nonzero (mod $p$) solution $(x_1, \ldots, x_{2p-1})$ to $f \equiv 0 \pmod p$. Fix such a solution.

Because $b$ is a quadratic nonresidue modulo $p$, the only solution to the equation $y_1^2 - by_2^2 \equiv 0 \pmod p$ is $y_1 \equiv y_2 \equiv 0 \pmod p$. It follows that, for the nonzero

(mod $p$) solution above, we must have

$$\sum_{i=1}^{2p-1} x_i^{p-1} \equiv 0 \pmod p \tag{1}$$

$$\sum_{i=1}^{2p-1} a_i x_i^{p-1} \equiv 0 \pmod p. \tag{2}$$

By Fermat's little theorem, if $x_i \not\equiv 0 \pmod p$, then $x_i^{p-1} \equiv 1 \pmod p$. If $I$ is the set of indices $i$ such that $x_i \not\equiv 0 \pmod p$, then Eq. (1) above becomes $\Sigma_{i \in I} 1 \equiv 0 \pmod p$. Since $0 < |I| \leqslant 2p - 1$, it follows that $|I| = p$.

Looking now at Eq. (2) above, we see that $\Sigma_{i \in I} a_i \equiv 0 \pmod p$, and thus we have found the required $n = p$ integers in this case.

We now proceed by induction on $n$. The result is true from the above work for $n = 2, 3$. Suppose $n \geqslant 4$ and that Theorem 1 holds for all smaller values of $n$. If $n$ is prime, then we are done by the first part, so we assume $n$ is composite, say $n = km$, where $1 < k, m < n$. Thus, Theorem 1 is true when $n$ is replaced by either $k$ or $m$.

An easy induction on $r \geqslant 2$ proves that if we have $rk - 1$ integers, $a_1, \dots, a_{rk-1}$, then there are $r - 1$ disjoint subsets $I_1, \dots, I_{r-1}$ of $\{1, 2, \dots, rk - 1\}$, each one of size $k$, such that $\Sigma_{i \in I_j} a_i \equiv 0 \pmod k$, $j = 1, 2, \dots, r - 1$. In particular, we use this fact with $r = 2m$. Thus, with our original $2n - 1 = (2m)k - 1$ integers, we have shown there are $2m - 1$ disjoint subsets $I_1, \dots, I_{2m-1}$ of $\{1, 2, \dots, 2n - 1\}$ such that, for $j = 1, \dots, 2m - 1$, $\Sigma_{i \in I_j} a_i \equiv 0 \pmod k$ and $|I_j| = k$.

For $j = 1, \dots, 2m - 1$, let $b_j = \Sigma_{i \in I_j} a_i$ and let $b_j' = b_j / k$; note that $b_j'$ is an integer. Now consider the $2m - 1$ integers $b_1', \dots, b_{2m-1}'$. By the inductive assumption, there is a subset $J$ of $\{1, 2, \dots, 2m - 1\}$ such that $|J| = m$ and $\Sigma_{j \in J} b_j' \equiv 0 \pmod m$. Let $I = \bigcup_{j \in J} I_j$. Evidently, $|I| = mk = n$ and $\Sigma_{i \in I} a_i \equiv \Sigma_{j \in J} \Sigma_{i \in I_j} a_i \equiv \Sigma_{j \in J} k b_j' \equiv 0 \pmod{km}$. Since $n = km$, we are done. ■

Chevalley's Theorem can be extended to systems of polynomial equations.

THEOREM 4. *For $j = 1, \dots, k$, let $F_j(x_1, \dots, x_n)$ be a polynomial of degree $r_j$ such that $F_j(0, \dots, 0) \equiv 0 \pmod p$, for some fixed prime $p$. If $\Sigma_j r_j < n$, then there is a nonzero* (mod $p$) *solution to the system of congruences $F_j \equiv 0 \pmod p$.*

Theorem 4 can be deduced from Theorem 3 by a suitable choice of polynomial $f(z_1, \dots, z_k)$. For we can then let $g(x_1, \dots, x_n) = f(F_1, \dots, F_k)$ and the equation $g \equiv 0 \pmod p$ is equivalent to the system of Theorem 4. For our needs in Theorem 1, we had $k = 2$ and $f(x, y) = x^2 - by^2$. In general, $f$ is fairly complicated, but always exists.

Observe that Theorem 4 may be used to prove Conjecture 2 in the event that $n$ is a prime. A result analogous to Theorem 4 for prime power was used in **[1]** to prove Conjecture 2 when $n$ is a prime power.

Since this article was written, Dom de Caen has informed us of an earlier proof of Theorem 1 in the literature. In **[4]** is our same 'multiplication' portion together with an elementary proof in the case $n$ is prime. Their techniques extend to the analogous result for any abelian group.

## REFERENCES

1. N. Alon, S. Friedland, and G. Kalai, Regular subgroups of almost regular graphs, *J. Combin. Theory*, Ser. B 37 (1984) 79–91.
2. Z. I. Borevich and I. R. Shafarevich, Number Theory, Academic Press, New York, 1966.

3. K. Ireland and M. Rosen, A Classical Introduction to Modern Number Theory, Springer-Verlag, New York, 1982.
4. P. Erdös, A. Ginzburg, and A. Ziv, A theorem in additive number theory, *Israel (State) Research and Development National Council Bulletin*, Section F, 10 (1961) 41–43.

# On Open Maps

JOHN CROWE and DOMINICK SAMPERI
*Department of Mathematics and Computer Science, Manhattan College, Bronx, NY 10471*

In this note we prove that if two topological spaces $X$ and $Y$ satisfy suitable hypotheses, and if $A$ is a subset of $X$ consisting of isolated points, then a continuous map $f\colon X \to Y$ is open if and only if $f \restriction X - A$ is an open map.

Our result (Theorem 1) applies when $X = \mathbb{R}^m$ and $Y = \mathbb{R}^n$, provided $n > 1$. Using the Implicit Function Theorem, it then follows that if $f\colon \mathbb{R}^m \to \mathbb{R}^n$ is $C^1$, if $n > 1$, and if $\{x \in \mathbb{R}^m \mid \operatorname{rank}(df(x)) < n\}$ consists of isolated points, then $f$ is an open map. This shows, in particular, that the Open Mapping Theorem for a non-constant analytic function $f(z)$ is a consequence of the Cauchy-Riemann equations and the fact that $f'(z) = 0$ only at isolated points.

Recall that a point $p$ in a connected subset $N$ of a topological space $X$ is called a cut point if $N - \{p\}$ is not connected.

We will say that a map $f\colon X \to Y$ is open at $x \in X$ if for any open neighborhood $N$ of $x$, $f(N)$ contains an open neighborhood of $f(x)$. Clearly, a map $f\colon X \to Y$ is open if and only if it is open at $x$ for each $x \in X$.

The following lemma will be useful.

LEMMA 1. *Let $Y$ be a locally connected topological space. Then the following two statements are equivalent.*

(i) *No connected open subset of $Y$ contains a cut point.*

(ii) *Whenever $p$ is an isolated point in the boundary of an open set $W \subset Y$, it follows that $p \in \overline{W}^0$.*

*Proof.* Assume that (ii) is false. Then there is a point $p$ in the boundary of an open set $W \subset Y$, with $p \notin \overline{W}^0$, and such that $(\overline{W} - W) \cap V = \{p\}$, for some open connected neighborhood $V$ of $p$. It then follows that $V \cap W \neq \varnothing$, $V \cap (\overline{W})^c \neq \varnothing$, and $V = (V \cap W) \cup (V \cap \overline{W}^c) \cup \{p\}$. Therefore, $p$ is a cut point in $V$, so (i) is false, and we have (i) $\Rightarrow$ (ii).

Now assume that (i) is false. Then there is a connected open set $N \subset Y$, with $N = W_1 \cup W_2 \cup \{p\}$, where $W_1$ and $W_2$ are disjoint open sets, and $p \notin W_1 \cup W_2$. It follows easily from this that $(\overline{W_1} - W_1) \cap N = \{p\}$, and $p \notin \overline{W_1}^0$, so (ii) is false. This concludes the proof of Lemma 1.

We are now ready to present our main result.

THEOREM 1. *Let $X$ be a locally compact Hausdorff space with the property that points in $X$ are not open subsets of $X$. Let $Y$ be a locally connected Hausdorff space*

*with the property that no open connected subset of $Y$ contains a cut point. Let $A$ be a subset of $X$ consisting of isolated points. Assume that the map $f$: $X \to Y$ is continuous, and that $f \restriction X - A$ is an open map.*

*Then $f$ is an open map.*

*Proof.* It will suffice to show that $f$ is open at $x$ for each $x \in A$. Assume that there is an $x_0 \in A$ such that $f$ is not open at $x_0$. Then there are open sets $N$ and $V$, with $\overline{N}$ compact, such that $x_0 \in N \subset \overline{N} \subset V$, and $f(x_0) \notin f(V)^0$. Since the points in $A$ are isolated, we may assume that $V \cap A = \{x_0\}$. Then since $f \restriction X - A$ is an open map, $f(V - \{x_0\})$ is an open subset of $f(V)$, so $f(x_0) \notin f(V - \{x_0\})$.

Since $\overline{N} - \{x_0\} \subset V - \{x_0\}$, $f(x_0) \notin f(\overline{N} - \{x_0\})$, and $f(N - \{x_0\})$ is open. Since $f \restriction \overline{N}$ is continuous and $\{x_0\}$ is not an open subset of $\overline{N}$, $f(x_0) \in \overline{f(N - \{x_0\})}$. On the other hand, $f(\overline{N})$ is closed, so

$$\overline{f(N - \{x_0\})} \subset f(N - \{x_0\}) \cup f(\overline{N} - N) \cup \{f(x_0)\},$$

and since $f(\overline{N} - N)$ is closed and does not contain $f(x_0)$, it follows that $f(x_0)$ is an isolated point in the boundary of the open set $f(N - \{x_0\})$. We have $f(x_0) \notin \overline{f(N - \{x_0\})}^0$, because $\overline{f(N - \{x_0\})} \subset f(\overline{N}) \subset f(V)$, and $f(x_0) \notin f(V)^0$. Therefore, (ii) in Lemma 1 is false for $Y$, so (i) must also be false, contradicting our hypothesis. This concludes the proof of Theorem 1.

REFERENCES

1. John L. Kelley, General Topology, D. Van Nostrand, Princeton, 1955.
2. Lynn Arthur Steen and J. Arthur Seebach, Jr., Counterexamples in Topology, Springer-Verlag, New York, 1970.

## On a Conjecture of R. J. Simpson About Exact Covering Congruences

DORON ZEILBERGER[1]
*Department of Mathematics, Drexel University, Philadelphia, PA 19104*

The following is a counterexample[2] to Simpson's conjecture [**2**]: $D = \{6, 15, 35, 14, 210 \text{ (140 times)}\}$. It was concocted using the elegant and powerful approach of [**1**].

REFERENCES

1. Marc A. Berger, Alexander Felzenbaum, and Aviezri S. Fraenkel, New results for covering systems of residue sets, *Bulletin (New Series) of the Amer. Math. Soc.*, 14 (1986) 121–125.
2. R. J. Simpson, Disjoint covering systems of congruences, this MONTHLY, 94 (1987) 865–868.

[1] Supported in part by NSF grant DMS 8800663.

[2] Another counterexample was found later, and independently, by John Beebee.

# THE TEACHING OF MATHEMATICS

EDITED BY MELVIN HENDRIKSEN AND STAN WAGON

## A Note on Taylor's Theorem

JOSÉ A. FACENDA AGUIRRE
*Facultad de Matemáticas, Tarfia s/n, 41012 Sevilla, Spain*

In a first course of mathematical analysis, when we study Taylor's formula, it is usual to prove the following approximation theorem on Taylor's polynomials:

*Let $f$: $[a, b] \to \mathbb{R}$ be a function $n$ times differentiable at $c \in [a, b]$. If*

$$P_n(h) = \sum_{k=0}^{n} \frac{f^{(k)}(c)}{k!} h^k,$$

then

$$\lim_{h \to 0} \frac{f(c+h) - P_n(h)}{h^n} = 0.$$

The proof is based on the application of L'Hôpital's rule $n-1$ times and the existence of $f^{(n)}(c)$; (see, e.g., [**5**, p. 193]).

The following result is a generalization of the theorem already mentioned and, furthermore, it naturally comes along with the study of functions of several variables.

THEOREM. *Let $A$ be an open subset of $\mathbb{R}^n$, and let $f$: $A \subset \mathbb{R}^n \to \mathbb{R}^m$ be a function $p$ times differentiable at $a \in A$. Then*

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - Df(a)(h) \cdots \frac{1}{p!} D^p f(a)(h^p)}{\|h\|^p} = 0.$$

Although the result is well known within the field of normed linear spaces (see [**4**, p. 165, Thm. 2.11.1]), the results used in the proof are not likely to be at the level of undergraduate students. Furthermore, we have not been able to find a published proof with respect to the aforementioned finite-dimensional spaces.

The purpose of this note is to give a short proof using only the mean value theorem for real-valued functions defined on any subset of $\mathbb{R}^n$. We think that it is interesting because of its simplicity and its being widely unknown.

Before beginning the proof we recall two results about higher-order derivatives.

(a) If $f$: $A \subset \mathbb{R}^n \to \mathbb{R}^m$ is $p$ times differentiable at $a \in A$, then $D^p f(a)$ is a $p$-linear symmetric form, that is, $D^p f(a) \in L_s^p(\mathbb{R}^n, \mathbb{R}^m)$ and

$$D^p f(a)(h^1, \dots, h^p) = \sum_{i_1=1}^{n} \cdots \sum_{i_p=1}^{n} D_{i_p \cdots i_1} f(a) h_{i_1}^1 \cdots h_{i_p}^p,$$

where $h^i = (h_1^i, \ldots, h_n^i) \in \mathbb{R}^n$, $1 \leqslant i \leqslant p$. If $h^1 = \cdots = h^p = h$, we denote

$$D^p f(a) \underbrace{(h, \ldots, h)}_{p \text{ terms}} = D^p f(a)(h^p).$$

If $\{j_1, \ldots, j_p\}$ is a permutation of $\{i_1, \ldots, i_p\}$, then

$$D_{i_p \cdots i_1} f(a) = D_{j_p \cdots j_1} f(a).$$

(See [**2**, p. 240, Thm. 7.83] for $p = 2$. For general $p$, see [**3**, 8.12.4 and 8.12.7]).

(b) Let $f$: $A \subset \mathbb{R}^n \to \mathbb{R}^m$ be a function $p$ times differentiable at $a \in A$. If we define $g(h) = D^p f(a)(h^p)$ then $g$ is a $C^\infty$-function and its successive derivatives are:

$$Dg(h) = pD^p f(a)(h^{p-1})$$

meaning

$$Dg(h)(k) = pD'f(a) \underbrace{(h, \ldots, h}_{p-1 \text{ terms}}, k), k \in \mathbb{R}^n.$$

$$D^2 g(h) = pD^p f(a)(h^{p-2})$$

meaning

$$D^2 g(h)(k^1, k^2) = p(p-1)D^p f(a) \underbrace{(h, \ldots, h}_{p-2 \text{ terms}}, k^1, k^2); \qquad k^1, k^2 \in \mathbb{R}^n.$$

And, in general, if $1 \leqslant i \leqslant p$ then

$$D^i g(h) = p(p-1) \cdots (p-i+1) D^p f(a)(h^{p-i})$$

meaning

$$D^i g(h)(k^1, \ldots, k^i)$$
$$= p(p-1) \cdots (p-i+1) D^p f(a) \underbrace{(h, \ldots, h}_{p-i \text{ terms}}, k^1, \ldots, k^i);$$

$k^1, \ldots, k^i \in \mathbb{R}^n$.

(See [**1**, p. 373] for the case $p = 2$.)

*Proof of the theorem.* We first suppose that $m = 1$ and proceed by induction on $p$. The result is true for $p = 1$ because $f$ is differentiable at $a$. We assume that the theorem is true for all $p \leqslant q$.

Let $f$ be $q+1$ times differentiable at $a \in A$. Then there exists $r > 0$ such that $B(a, r) \subset A$, and for every $x \in B(a, r)$, $D^q f(x)$ there exists. We define, for $h \in B(0, r)$,

$$R(h) = f(a+h) - f(a) - Df(a)(h) - \cdots - \frac{1}{(q+1)!} D^{q+1} f(a)(h^{q+1}).$$

It is obvious that $R$ is $q+1$ times differentiable at 0, so its partial derivatives, $D_i R = R^i$, $1 \leqslant i \leqslant n$, are $q$ times differentiable at the origin. Hence, by the

inductive hypothesis, if

$$S^i(h) = R^i(h) - R^i(0) - DR^i(0)(h) - \cdots - \frac{1}{q!}D^qR^i(0)(h^q),$$

it follows that

$$\lim_{h\to 0}\frac{S^i(h)}{\|h\|^q} = 0, \qquad 1 \leqslant i \leqslant n.$$

We claim that $S^i(h) = R^i(h)$, $1 \leqslant i \leqslant n$.

To obtain this it is sufficient to prove $D^jR(0) = 0$, $1 \leqslant j \leqslant q + 1$, since, according to the expression of higher-order derivatives given in $(A)$, it follows that $D^jR^i(0) = 0$, $0 \leqslant j \leqslant q$ and $1 \leqslant i \leqslant n$.

With this aim, we differentiate $R$ repeatedly, obtaining

$$D^jR(h) = D^jf(a+h) - D^jf(a) - \cdots - \frac{1}{(q-j+1)!}D^{q+1}f(a)(h^{q-j+1}),$$

$1 \leqslant j \leqslant q$. In particular,

$$D^qR(h) = D^qf(a+h) - D^qf(a) - D^{q+1}f(a)(h).$$

So

$$D^jR(0) = 0 \in L_s^j(\mathbb{R}^n, \mathbb{R}), \qquad 1 \leqslant j \leqslant q.$$

(Note that result (b) is being used.)

Moreover, since

$$\lim_{h\to 0}\frac{D^qR(h)}{\|h\|} = \lim_{h\to 0}\frac{D^qf(a+h) - D^qf(a) - D^{q+1}f(a)(h)}{\|h\|} = 0$$

because $f$ is $q+1$ times differentiable at $a$, it follows that $D^{q+1}R(0) = 0 \in L_s^{q+1}(\mathbb{R}^n, \mathbb{R})$. Thus $S^i(h) = R^i(h)$, as we stated.

Therefore, we have

$$\lim_{h\to 0}\frac{D_iR(h)}{\|h\|^q} = 0, \qquad 1 \leqslant i \leqslant n.$$

If we apply the mean value theorem we infer that

$$R(h) = R(h) - R(0) = DR(th)h = \sum_{i=1}^{n} D_iR(th)h_i, \qquad 0 < t < 1.$$

We conclude the proof from the inequalities

$$\frac{|R(h)|}{\|h\|^{q+1}} \leqslant \sum_{i=1}^{n}\frac{|D_iR(th)|}{\|h\|^q}\frac{|h_i|}{\|h\|} \leqslant \sum_{i=1}^{n}\frac{|D_iR(th)|}{\|h\|^q}$$

in which the last term tends towards 0 as $h$ does.

In the case when $f$ takes values in $\mathbb{R}^m$, $m > 1$, we apply the same argument to each of the real-valued functions $f_i$, $1 \leqslant i \leqslant m$, which occur in the coordinate representation of the mapping $f$.

We conclude with the following remainder form of the theorem:

COROLLARY. *If $f$ is $p$ times differentiable at $a$, then there exists a function $E_p$ defined on a neighborhood of* 0 *such that*

$$f(a+h)=f(a)+Df(a)(h)+\cdots+\frac{1}{p!}D^pf(a)(h^p)+E_p(h)\|h\|^p$$

and $E_p(h)\to 0$ as $h\to 0$.

The author thanks the referee for many valuable suggestions.

REFERENCES

1. R. G. Bartle, The Elements of Real Analysis, 2nd ed., Wiley, New York, 1976.
2. J. C. Burkill and H. Burkill, A Second Course in Mathematical Analysis, Cambridge Univ. Press, Cambridge, 1970.
3. J. Dieudonné, Foundations of Modern Analysis, Academic Press, New York, 1960.
4. M. J. Field, Differential Calculus and its Applications, Van Nostrand Reinhold, New York, 1976.
5. K. R. Stromberg, An Introduction to Classical Real Analysis, Wadsworth, Belmont, 1981.

# Material Implication Revisited

JOSEPH S. FULDA
*Department of Computer Science, Hofstra University, Hempstead, NY 11550*

Material implication, the logical connective corresponding to the English "if," has vexed scholars and students alike with its counterintuitive definition of conditionals with false antecedents as true (see Figure 1) [**1**]. In addition, using conditionals with unrelated antecedents and consequents, something which is not proscribed, seems to be contrary to the intuitive notion of implication. These two consequences of the traditional definition of material implication have led to the so-called paradoxes of material implication: $p\Rightarrow(q\Rightarrow p)$, a true proposition is implied by any proposition whatever; and $\sim p\Rightarrow(p\Rightarrow q)$, a false proposition implies any proposition whatever [**2**].

| $p$ | $q$ | $\sim p\vee q$ |
|---|---|---|
| T | T | T |
| T | F | F |
| F | T | T |
| F | F | T |

FIG. 1. The definition of material implication, $p\Rightarrow q$.

We will show in this paper that the traditional definition of material implication, when properly understood, is quite intuitive. We need not revert to a three-valued logic, to modality, to alternative, non-truth functional semantics for the conditional, or to the extensive literature on counterfactuals. These developments of modern symbolic logic are of major importance. But to explain the simple material conditional, we need not go beyond the framework of the classical, two-valued, truth-functional logic.

First, I remind the reader that if we work within the formalism of the classical, two-valued, truth-functional logic, the standard truth table for the conditional must result. All alternative truth tables that might be used for material implication clearly result in some other connective, such as conjunction, equivalence, or merely the conditional's consequent (see FIG. 3). Yet this demonstration that the standard truth table for the conditional is the only reasonable alternative within the classical system of two-valued, truth-functional logic does not suffice as an *explanation* of the definition of $\Rightarrow$. As philosopher Robert Nozick writes: "A proof of $p$ will give us the conviction that $p$ is true, but it need not give us understanding of how $p$ *can* be true, given apparent excluders" [**3**] (slightly edited).

One explanation often offered is that of C. Ray Wylie [**4**]. Wylie reminds us that we often use implications with false antecedents in natural language and gives examples of this. These examples provide further conviction that the traditional definition of material implication is appropriate, but still leave considerable uneasiness with it.

It is the thesis of this paper that this uneasiness is none other than the familiar temptation to commit the fallacy of conversion ($p \Rightarrow q \vdash q \Rightarrow p$), also known as the fallacy of affirming the consequent ($p \Rightarrow q, q \vdash p$) [**5**], and replace material implication ($\Rightarrow$) with material equivalence ($\Leftrightarrow$).

One reason for this temptation is that "if" as used in everyday language often means iff—if and only if. For example, two of the four types of implication that Copi isolates in his much-used text [**6**] are really biconditionals. One is definitional implication: definitions must be reversible. Thus the converse of "If a polygon has three sides, then it is a triangle" is true also. Another is causal implication. If $a$ then $b$, because $a$ causes $b$; therefore, if $b$ then $a$, since for $b$ to have occurred, its cause, $a$, must have first occurred. (We sidestep here multiple causation and a vast literature on causality and how to treat it logically.) Thus when a child is told by a parent, "If you don't behave, I shall slap you," "if" is used in the sense of iff and the child is right to consider good behavior adequate protection from being slapped. Were this not so, there would be no particular reason for good behavior and hence no particular reason for the parent's utterance.

In Latin, there is a grammatical difference between *if* used as $\Rightarrow$ and *if* used as $\Leftrightarrow$ [**7**]. Thus, *si laborabit*, *felix erit*, if he works he will be happy, is future indicative in both the antecedent and the consequent. But, *si laboraverit*, *felix erit*, if he works (will have worked) he will be happy, is future perfect indicative in the antecedent and future simple indicative in the consequent. As Moreland and Fleischer explain, "In such cases it is emphasized that the action in the protasis *must* be completed in order for the action in the apodosis to occur" (italics in original) [**7**].

However, there is a deeper reason for the desire to use $\Leftrightarrow$ in place of $\Rightarrow$. Not only is $\Leftrightarrow$ given by the standard definitions of material equivalence as $(p \Rightarrow q) \wedge (q \Rightarrow p)$ and as $(p \wedge q) \vee (\sim p \wedge \sim q)$, but also by $\sim p \oplus q$ as shown in Figure 2. Observe that this definition of material equivalence differs from the traditional

| $p$ | $q$ | $\sim p \oplus q$ |
|---|---|---|
| T | T | T |
| T | F | F |
| F | T | F |
| F | F | T |

FIG. 2. The definition of material equivalence, $p \Leftrightarrow q$.

definition of material implication only in that the disjunction is exclusive rather than inclusive, a difference easily and naturally slipped around.

Every time we use "if" to mean iff in everyday language, we are using $\oplus$ for "or." Thus, in the example above, the parent is telling the child that either he must behave or he will be slapped, but of course not both.

Let us now turn to conditionals with unrelated antecedents and consequents. In a way, this problem is greater than the problem of conditionals with false antecedents, for the latter causes us trouble only with the last two lines of the standard truth table, while with the former only the second line of the truth table is unproblematic.

The substitution of $\oplus$ for $\vee$ that we have been discussing also prevents a prime way that unrelatedness between the antecedent and the consequent of a conditional can arise in the course of proving the validity of an argument. In particular, we observe that most of the rules of the various systems for natural deduction apply to $\oplus$ as well as to $\vee$ (e.g., commutation, association, distribution of conjunction over disjunction, disjunctive syllogism, constructive dilemma, destructive dilemma). One clear and notable exception is the rule often called *addition*: from $p$ we may infer $p \vee q$, but we cannot infer $p \oplus q$, because $q$ might be true. Indeed, for exclusive disjunction, the following rule, reminiscent of disjunctive syllogism, holds: $p \oplus q, p \vdash \sim q$. Not only is addition a rule with no $\oplus$-analogue, it is also the only rule in all natural deduction systems known to the present writer in which an irrelevancy is introduced. A term nowhere in the premises occurs in the conclusion. Thus by replacing $\vee$ by $\oplus$ in the definition of material implication (and obtaining material equivalence), one is ruling out the irrelevancy stemming from $\vee$. Such irrelevancy can show up in conditionals with unrelated antecedents and consequents, which is yet another reason why replacing $\vee$ with $\oplus$ and hence $\Rightarrow$ with $\Leftrightarrow$ seems so natural and comfortable a thing to do.

Observe that the paradoxes of material implication are also made possible by addition, as shown by the following conditional proofs (after Copi's method [**8**]):

| $\vdash\ p \Rightarrow (q \Rightarrow p)$ | |
|---|---|
| 1. $p$ | assumption |
| 2. $p \vee \sim q$ | addition, 1 |
| 3. $\sim q \vee p$ | commutation, 2 |
| 4. $q \Rightarrow p$ | def. mat. impl., 3 |
| 5. $p \Rightarrow (q \Rightarrow p)$ | C. P., 1, 2–4 |

| $\vdash\ \sim p \Rightarrow (p \Rightarrow q)$ | |
|---|---|
| 1. $\sim p$ | assumption |
| 2. $\sim p \vee q$ | addition, 1 |
| 3. $p \Rightarrow q$ | def. mat. impl., 2 |
| 4. $\sim p \Rightarrow (p \Rightarrow q)$ | C. P., 1, 2–3 |

The absence of these paradoxes when $\oplus$ is used rather than $\vee$ is a final reason for the temptation to replace $\Rightarrow$ by $\Leftrightarrow$.

Finally, it is interesting to note that only $\Rightarrow_1$ and $\Rightarrow_3$ of the four possibilities for the truth table of material implication (see FIGURE 3) leave $p \Rightarrow q \wedge q \Rightarrow p$ undisturbed as the definition of p $\Leftrightarrow q$. This may be yet another reason for the temptation to substitute $\Rightarrow_3$ for $\Rightarrow_1$, while there is no similar temptation to

| $p$ | $q$ | $p \Rightarrow_1 q$ | $p \Rightarrow_2 q$ | $p \Rightarrow_3 q$ | $p \Rightarrow_4 q$ |
|---|---|---|---|---|---|
| T | T | T | T | T | T |
| T | F | F | F | F | F |
| F | T | T | T | F | F |
| F | F | T | F | T | F |

FIG. 3. Four alternatives for $\Rightarrow$.

substitute either $\Rightarrow_2$ or $\Rightarrow_4$ for $\Rightarrow_1$. Indeed, while $p \Rightarrow_1 q \wedge q \Rightarrow_1 p$ and $p \Rightarrow_3 q \wedge q \Rightarrow_3 p$ both yield *iff*, both $p \Rightarrow_2 q \wedge q \Rightarrow_2 p$ and $p \Rightarrow_4 q \wedge q \Rightarrow_4 p$ yield *and*. This can be easily verified by constructing the appropriate truth tables.

In summary: (1) The standard truth table for material implication obtained within the framework of classical, two-valued, truth-functional logic is the only possibility that does not clearly mean something other than implication. (2) All truth table entries in the definition of material implication correspond to uses in everyday speech and writing, including the entries with false antecedents. (3) Remaining uneasiness with the definition of material implication can be attributed to a desire to commit the fallacy of conversion and replace $\Rightarrow$ by $\Leftrightarrow$. (4) This fallacy is made attractive by everyday use of "if" to mean iff. (5) Such a substitution amounts to using $\oplus$ in place of $\vee$ in the definition of material implication, something easily and naturally done. (6) The use of $\oplus$ in place of $\vee$ also removes any unrelatedness between the antecedent and the consequent of a conditional introduced by the addition rule which can be used with $\vee$ but not with $\oplus$. (7) Their proofs indicate that the paradoxes of material implication are likewise made possible by addition and are thus eliminated when $\oplus$, rather than $\vee$, is used.

REFERENCES

1. See, for example, G. Bedell, Teaching the material conditional, *Teaching Philosophy*, 2 (1977–78) 225–236. Bedell's approach is quite different from ours.
2. I. Copi, *Symbolic Logic*, 4th edition, Macmillan, New York, 1973, p. 26.
3. R. Nozick, *Philosophical Explanations*, Harvard University Press, Cambridge, 1981, p. 10.
4. C. R. Wylie, 'False implies false' is true, *The Mathematics Teacher*, 72 (1979) 404–405. See also J. Bookman, Why 'false $\rightarrow$ false' is true—a discovery explanation, *The Mathematics Teacher*, 71 (1978) 675–676 and the correspondence thereon in *The Mathematics Teacher*, 72 (1979) 405. See also I. Copi, op. cit., p. 16.
5. I. Copi, op. cit., p. 22.
6. ______, pp. 14–16.
7. F. Moreland and R. Fleischer, Latin: An Intensive Course, University of California Press, Berkeley, 1977, "Conditional Sentences," pp. 38–39.
8. I. Copi, op. cit., pp. 58–61.

## A Pictorial Proof of Uniform Continuity

D. M. Bloom
*Mathematics Department, Brooklyn College of CUNY, Brooklyn, NY 11210*

The following theorem is usually proved in a beginning analysis course.

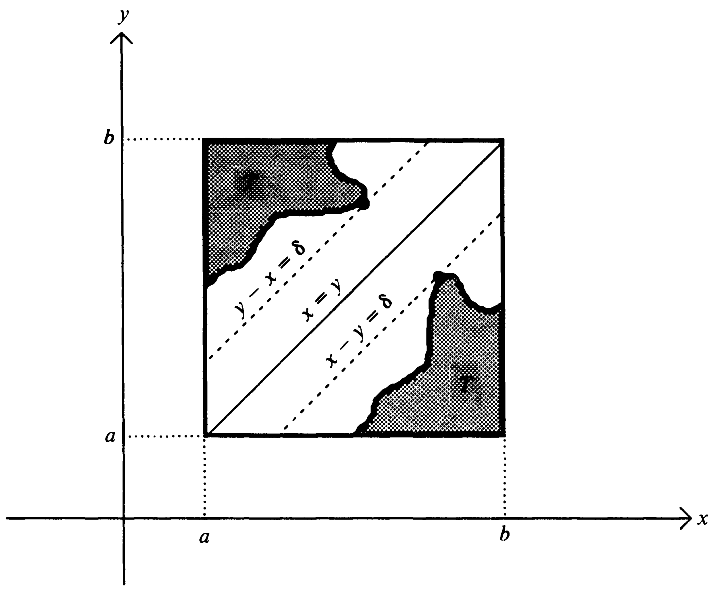
THEOREM. *If the real-valued function $f$ is continuous on the interval $[a, b]$ then $f$ is uniformly continuous on $[a, b]$.*

The standard proofs use either infinite sequences or (if one has discussed compactness) open coverings (see [**1**, p. 103] or [**2**, pp. 78–79]). Here is a proof which uses neither; instead, we use facts about closed sets and continuity in the

plane. (Admittedly, these facts must be proved, but one has usually done so anyway.) The idea is that if $f$ is continuous on $[a, b]$ and if $\varepsilon > 0$, then the $\delta$ which will work for the given $\varepsilon$ is simply the minimum value of $|x - y|$ on the closed set

$$T = \{(x, y): x, y \in [a, b] \text{ and } |f(x) - f(y)| \geqslant \varepsilon\},$$

as illustrated by the following diagram:




$T = \{(x, y): x, y \in [a, b]$ and $|f(x) - f(y)| \geqslant \varepsilon\}$

Notice that our $\delta$ must in fact be positive (not 0), since if $T$ intersects the line $x = y$ then for some $x$ we have $|f(x) - f(x)| \geqslant \varepsilon$, which is absurd.

To fill in more of the reasoning: since $f$ is continuous on $I = [a, b]$, it follows that $g(x, y) = |f(x) - f(y)|$ is continuous on $I \times I$; hence the set $T = g^{-1}[\varepsilon, \infty)$ is closed. Since $T$ is also bounded, and since the distance function $d(x, y) = |x - y|$ is continuous on $\mathbb{R}^2$, it follows that the number $\delta = \min\{|x - y|: (x, y) \in T\}$ exists. As shown above, $\delta > 0$; moreover,

$$|x - y| < \delta \Rightarrow (x, y) \notin T \Rightarrow g(x, y) < \varepsilon \Rightarrow |f(x) - f(y)| < \varepsilon,$$

as desired.

This proof easily generalizes to compact metric spaces. The facts that are needed for the proof are (1) the inverse image of a closed set under a continuous function is closed, (2) a closed subset of a compact space is compact, (3) if $C$ is compact then so is $C \times C$, (4) a continuous function from a compact space to $\mathbb{R}$ has a minimum value, and (5) the distance function $d(x, y)$ is always continuous.

## REFERENCES

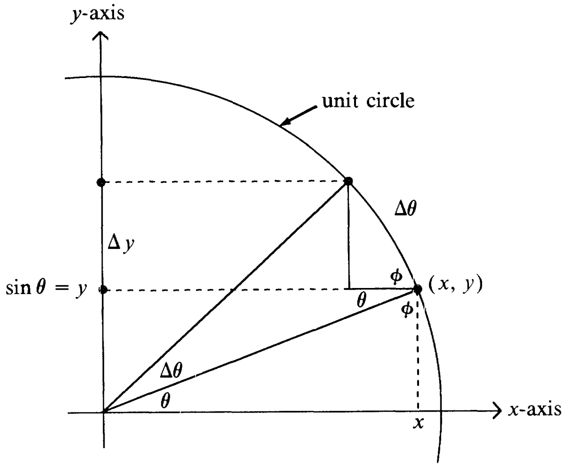

1. K. Ross, Elementary Analysis: The Theory of Calculus, Springer-Verlag, New York, 1980.
2. W. Rudin, Principles of Mathematical Analysis, 2nd ed., McGraw-Hill, New York, 1964.

# On the Differentiation Formula for sin $\theta$

Donald Hartig
*Mathematics Department, California Polytechnic State University, San Luis Obispo, CA 93407*

Very few textbooks on elementary calculus can lay claim to a "rigorous" proof of the fact that the sine function differentiates to the cosine function. See the remarks in Peter Ungar's insightful review [**1**] of three such texts. In the figure below I offer what should be regarded as another plausibility argument in support of the fact that

$$\frac{d}{d\theta}\sin\theta = \cos\theta.$$




$$\frac{dy}{d\theta} \approx \frac{\Delta y}{\Delta\theta} \approx \sin\phi = \cos\theta.$$

It is as rigorous as the usual "proofs" and has the redeeming feature that the companion formula:

$$\frac{d}{d\theta}\cos\theta = -\sin\theta,$$

can be motivated in the same way:

REFERENCES

1. Peter Ungar, Review of *Calculus and Analytic Geometry* by Al Shenk, *Calculus with Analytic Geometry* by M. A. Munem and D. J. Foulis, and *Calculus with Analytic Geometry* by Howard Anton, this Monthly, 93 (1986) 221–230.

E 3316. *Proposed by Jim Delany, California Polytechnic State University, San Luis Obispo.*

Consider the fluctuations in distance between an observer located at $(h, k)$ and an object orbiting along the ellipse $x^2/a^2 + y^2/b^2 = 1$, where $a > b > 0$. If the observer is at or near the origin, then on each orbit the distance will have two relative maxima and two relative minima. If the observer is far from the origin, there will be only one of each. Determine the equation of the curve that marks the boundary between the two cases.

E 3317. *Proposed by Peter J. Ferraro, Roselle Park, NJ.*

For any positive integer $n$ and any positive real number $\alpha$ put

$$S_n(\alpha) = \sum_{j=0}^{n-1} \lfloor \sqrt{\alpha + j/n} \rfloor .$$

(a) Prove that $S_n(\alpha) = n\lfloor \sqrt{\alpha} \rfloor$ if $\lfloor \alpha \rfloor + 1$ is not a perfect square.
(b) What is the value of $S_n(\alpha)$ if $\lfloor \alpha \rfloor + 1$ is a perfect square?

Here $\lfloor x \rfloor$ denotes the greatest integer not exceeding $x$.

E 3318. *Proposed by Walter Rudin, University of Wisconsin, Madison.*

Suppose $0 < p < \infty$. Let $A$ be the set of sequences of positive real numbers $a_1, a_2, \ldots$ such that $\Sigma_{n=1}^{\infty} a_n = 1$. Find $\sup_A \Sigma_{n=1}^{\infty} a_n(a_n + a_{n+1} + \cdots)^{p-1}$ and $\sup_A \Sigma_{n=1}^{\infty} a_n(a_1 + a_2 + \cdots + a_n)^{p-1}$.

## SOLUTIONS OF ELEMENTARY PROBLEMS

### Weighing Seven Coins in Five Weighings

E 3023 [1983, 645]. *Proposed by J. G. Mauldon, Amherst College, Amherst, MA.*

You are given an accurate indicating spring scale (not a balance) and seven coins each of which weighs $x$ or $y$, where $x$ and $y$ are unknown. In five weighings, determine the weight of each coin.

*Solution I by Aage Bondesen.* We assume $x \leqslant y$. We let $a, b, c, d, e, f, g$ denote both the coins and their weights, and we use a similar notation for the sets of coins to be defined below and the weights of those sets. We use our first two weighings to weigh the coin sets $A = \{a, b, c\}$ and $B = \{a, d, e\}$. Our argument will be much concerned with the "average weights" $\alpha = A/3$ and $\beta = B/3$. By symmetry, we may assume $\alpha \leqslant \beta$. We treat separately the cases $\alpha < \beta$ and $\alpha = \beta$.

I) $\alpha < \beta$. We spend our third weighing on the coin set $C = \{b, d\}$. We put $\gamma = C/2$ and divide into five subcases. Each time we compute a ratio of differences, the denominator is 0 if and only if $x = y$, in which case $x = y = \alpha$ and we know all the weights, so we may assume $x < y$.

1) $\gamma < \alpha < \beta$. This implies (1) $b + 3d < 2a + 2c$ and (2) $b + c < e + d$. Consider the possibilities for $d$. If $d = x$, then (2) implies $b = c = x$ and $e = y$, and

then (1) yields $a = y$. If $d = y$, then (1) implies $a = c = y$ and $b = x$, and then (2) yields $e = y$. These possibilities can be distinguished by $\alpha, \beta, \gamma$, because in the first case $(\alpha - \gamma)/(\beta - \alpha) = (y - x)/(y - x) = 1$, while in the second $(\alpha - \gamma)/(\beta - \alpha) = (y - x)/2(y - x) = 1/2$. Having distinguished these, we can find $x, y$ from $2x + y = A$ and $x + 2y = B$ in the first case and from $x + 2y = A$ and $3y = B$ in the second case. Now $a, b, c, d, e$ are known, and $f, g$ can be determined in a fourth and fifth weighing.

2) $\gamma = \alpha < \beta$. $\gamma = \alpha$ implies $a = b = c = d = \alpha$, and $\alpha < \beta$ implies $\alpha = x$ and $e = y$. Now $f, g$ can be found in two more weighings.

3) $\alpha < \gamma < \beta$. This implies (1) $2a + 2c < b + 3d$ and (2) $3b + d < 2a + 2e$. We will consider $(\gamma - \alpha)/(\beta - \gamma)$ under the two possibilities $a = x$ and $a = y$. If $a = x$, then (2) implies $b = x$ and $e = y$, and then (1) yields $d = y$. Now $(\gamma - \alpha)/(\beta - \gamma) = (3y - x - 2c)/(y - x)$, which is 3 for $c = x$ and 1 for $c = y$. On the other hand, if $a = y$, then (1) implies $d = y$ and $c = x$, and then (2) yields $b = x$. In this case $(\gamma - \alpha)/(\beta - \gamma) = (y - x)/(y + 2e - 3x)$, which is 1 if $e = x$ and $1/3$ if $e = y$.

We conclude that if $(\gamma - \alpha)/(\beta - \gamma)$ is 3 or $1/3$ we have $(a, b, c, d, e) = (x, x, x, y, y)$ or $(a, b, c, d, e) = (y, x, x, y, y)$, respectively. We then find $x, y$ from $3x = A$ and $x + 2y = B$ or $y + 2x = A$ and $3y = B$, respectively, which yields $a, b, c, d, e$, and we have two further weighings to determine $f, g$. If $(\gamma - \alpha)/(\beta - \gamma) = 1$, we must weigh one more set to decide between the two possibilities for $(a, b, c, d, e)$ in terms of $x, y$. We choose to weigh $D = \{c, e, f\}$ and put $\delta = D/3$. If $(a, b, c, d, e) = (x, x, y, y, y)$, then $(\alpha - \delta)/(\beta - \alpha) = (2x - y - f)/(y - x)$, which is $-1$ if $f = x$ and $-2$ if $f = y$. On the other hand, if $(a, b, c, d, e) = (y, x, x, y, x)$, then $(\alpha - \delta)/(\beta - \alpha) = (y - f)/(y - x)$, which is 1 if $f = x$ and 0 if $f = y$. Hence evaluating this ratio determines $a, b, c, d, e$ and also $f$ as $x$ or $y$. We find $x, y$ from $2x + y = A$ and $x + 2y = B$ in each case, and the fifth weighing determines $g$.

The other two cases, 4) $\alpha < \beta = \gamma$ and 5) $\alpha < \beta < \gamma$, are completely symmetric to 2) and 1), respectively. This reduces us to the other major case.

II) $\alpha = \beta$. We must once more define the coin sets $C, D, E$ to be weighed besides $A$ and $B$. We take $C = \{a, b, d, f\}$ and $D = \{a, c, d, f\}$ and put $\gamma = C/4$ and $\delta = D/4$. We may assume $\gamma \leqslant \delta$. We treat $\gamma < \delta$ and $\gamma = \delta$ separately.

II′) $\gamma < \delta$. In this case, $b = x$, $c = y$, and $x \neq y$. Since $\alpha \neq \gamma$ and $\alpha \neq \delta$ (otherwise $a = b = c = d = e = f$ and $\gamma = \delta$), we have three subcases.

1) $\alpha < \gamma < \delta$. This leads to $a + x + 4y < 3f + 3d$, whence $f = d = y$, and then $e = x$ (since $\alpha = \beta$). Now comparison of $3\gamma$ with $2\alpha + \delta$ reduces to comparison of $b + 6d + 6f$ with $2a + 11c$, which reduces to comparison of $x + y$ with $2a$. We have $a = x \Leftrightarrow 3\gamma > 2\alpha + \delta$ and $a = y \Leftrightarrow 3\gamma < 2\alpha + \delta$. Thus we know whether $a$ is $x$ or $y$, and we can then determine $x$ and $y$ from $A$ and $C$. Then $a, b, c, d, e, f$ are known, and $g$ can be determined in a fifth weighing.

2) $\gamma < \delta < \alpha$. This case is symmetric to 1).

3) $\gamma < \alpha < \delta$. This implies $3x + 3d + 3f < a + 4x + 4y < 3y + 3d + 3f$, i.e., $a + y + 4x < 3f + 3d < a + x + 4y$, whence $f + d = x + y$. Since also $e + d = x + y$ (from $\alpha = \beta$), it follows that $f = e$, and then comparison of $\alpha + \beta$ with $\gamma + \delta$ reduces to comparison of $2a$ with $x + y$. We have $a = x \Leftrightarrow \alpha + \beta < \gamma + \delta$ and $a = y \Leftrightarrow \alpha + \beta > \gamma + \delta$. We now weigh $E = \{e, f, g\}$ in our fifth weighing and put $\varepsilon = E/3$. If $a = x$, then $6(\delta - \alpha) = y - x$ and $6(\varepsilon - \delta) = 4e + 2g -$

$3x - 3y$. On the other hand, if $a = y$, then $6(\gamma - \alpha) = x - y$ and $6(\varepsilon - \gamma) = 4e + 2g - 3x - 3y$. In the two cases, we compute $(\delta - \alpha)/(\varepsilon - \delta)$ or $(\gamma - \alpha)/(\varepsilon - \gamma)$ to identify $e$ and $g$ from the following tables:

$a = x$

| $e$ | $x$ | $x$ | $y$ | $y$ |
|---|---|---|---|---|
| $g$ | $x$ | $y$ | $x$ | $y$ |
| $\dfrac{\delta - \alpha}{\varepsilon - \delta}$ | $-\frac{1}{3}$ | $-1$ | $1$ | $\frac{1}{3}$ |

$a = y$

| $e$ | $x$ | $x$ | $y$ | $y$ |
|---|---|---|---|---|
| $g$ | $x$ | $y$ | $x$ | $y$ |
| $\dfrac{\gamma - \alpha}{\varepsilon - \gamma}$ | $\frac{1}{3}$ | $1$ | $-1$ | $-\frac{1}{3}$ |

Once $e$ and $g$ are identified to $x$ or $y$, the same can be done with $f$ and $d$, $x$ and $y$ can be calculated from $A$ and $C$, and we are finished.

II″) $\gamma = \delta$. Since also $\alpha = \beta$, we have $b = c = d = e$. We divide into three subcases.

1) $\alpha = \gamma$. In this case $a = b = c = d = e = f = \alpha$, and $g$ is found in a fifth weighing.

2) $\alpha < \gamma$. This reduces to $a + 2b < 3f$, whence $f = y$, and $a$ or $b$ is $x$. We spend our fifth weighing on $E = \{a, g\}$ and put $\varepsilon = E/2$. The following table is easily established (note that $y \neq x$):

| $g$ | $x$ | $x$ | $x$ | $y$ | $y$ | $y$ |
|---|---|---|---|---|---|---|
| $a$ | $x$ | $y$ | $x$ | $x$ | $y$ | $x$ |
| $b$ | $y$ | $x$ | $x$ | $y$ | $x$ | $x$ |
| $\dfrac{\alpha - \varepsilon}{\gamma - \alpha}$ | $8$ | $-1$ | $0$ | $2$ | $-4$ | $-2$ |

This assigns $x$ or $y$ to each coin, and the values of $x$ and $y$ can be calculated from $A$ and $C$.

3) $\gamma < \alpha$. This case is symmetric to 2).

*Solution II by the proposer.* We give a strong (i.e., nonbranching) weighing procedure which is robust, in the sense that modest manufacturing, experimental or rounding errors will not give a qualitatively wrong result, unless perhaps when the coins are all of the same weight or the values of $x$ and $y$ are practically indistinguishable. Part of our solution may be regarded as a computer-assisted proof that five weighings suffice for the solution of the problem.

We denote the seven coins by $a, b, c, d, e, f, g$. Our procedure is to determine, in five weighings, the values of $P, Q, R, S, T$—the weights, respectively, of the five sets $\{c, d, e, f, g\}$, $\{a, b, c, d\}$, $\{a, d, e\}$, $\{b, e, f\}$ and $\{a, f, g\}$.

We assume without loss of generality that coin $a$ is of weight $y$ and we denote by $X$ the set of all coins of weight $x$, so that either $X = \varnothing$ or else $X \in \Sigma$, where $\Sigma$ denotes the set of all the 63 nonempty subsets of $\{b, c, d, e, f, g\}$. We write $p, q$ for the number of coins of weight $x$ in the two sets $\{c, d, e, f, g\}$, $\{a, b, c, d\}$ respectively, observing that $p$ is an integer and $q \in \{0, 1, 2, 3\}$, so that the relation $4p = 5q$ would imply $p = q = 0$ and $X = \varnothing$.

If $X = \varnothing$ all the coins have weight $y$ and so $4P = 5Q$. Conversely, if $4P = 5Q$, we have $(4p - 5q)(x - y) = 4P - 5Q = 0$ so that either $x = y$ or $X = \varnothing$, and in either case the seven coins are all of the same weight $P/5$ and the problem is solved.

Henceforth, therefore, we may and shall assume that $4P \neq 5Q$, so that $4p \neq 5q$, $x \neq y$ and $X \in \Sigma$.

Once the set $X \in \Sigma$ is known, then $p$ and $q$ are also known and we find that $y = (Qp - Pq)/(4p - 5q)$ and $x = y + (4P - 5Q)/(4p - 5q)$, so that the problem is solved. All that remains, therefore, is to determine the set $X$.

Compute the two quantities $U, V$, where $U = U(X, x, y) = 4P - 5Q \neq 0$ and $V = V(X, x, y) = 3Q - 3R - 21S + 20T$, observing that $U(X, x, y) = (x - y)U(X, 1, 0)$ and $V(X, x, y) = (x - y)V(X, 1, 0)$, so that $V(X, x, y)/U(X, x, y)$ is independent of $x$ and $y$. This observation justifies the notation $W = W(X)$ in the expression $W = W(X) = 9240(60 + V(X, x, y)/U(X, x, y))$ and permits us, in evaluating $W(X)$, to assume that $x = 1$ and $y = 0$. Then $U = 4p - 5q$, so that $U$ is an integer, and similarly $V$ is an integer and we find $1 \leqslant |U| \leqslant 12$ and $|U| \neq 9$, so that $U$ divides 9240 and consequently, for every $X \in \Sigma$, $W(X)$ is an integer.

| *abcdefg* | $W(X)$ | *abcdefg* | $W(X)$ | *abcdefg* | $W(X)$ | *abcdefg* | $W(X)$ |
|---|---|---|---|---|---|---|---|
| ······· | $0 \div 0$ | ···×××· | 521400 | ·····×· | 552090 | ··×··×× | 583440 |
| ·×××××· | 184800 | ·××·××× | 523600 | ··××××× | 552552 | ·×····· | 587664 |
| ·×·×××· | 355740 | ·×··××× | 524040 | ··×·××× | 552720 | ··××·×× | 588280 |
| ·××·××· | 369600 | ··×·××· | 525360 | ··×××·× | 552860 | ·×××·×× | 591360 |
| ·××××·× | 378840 | ····××· | 525525 | ··×·×·× | 553080 | ······× | 600600 |
| ·×··××· | 421960 | ··×···· | 526680 | ···×··· | 554400 | ·×××·×· | 603680 |
| ·×·××·× | 452760 | ·××···× | 531300 | ·×···×× | 557480 | ···×··× | 616000 |
| ··×××·· | 457380 | ·×····× | 535920 | ·×·×·×× | 559020 | ··×···× | 625240 |
| ·××·×·× | 466620 | ·×××··× | 539000 | ··×··×· | 560560 | ·××··×· | 628320 |
| ···××·· | 480480 | ··××··· | 540540 | ··××·×· | 563640 | ·×·×·×· | 642180 |
| ·×··×·× | 486640 | ·×·×··× | 545160 | ·××··×× | 572880 | ··××··× | 660660 |
| ··×·×·· | 489720 | ···××·× | 549120 | ·×××··· | 574200 | ·××××·· | 674520 |
| ····×·· | 498960 | ····×·× | 549780 | ·····×× | 576345 | ·×···×· | 729960 |
| ·×××××× | 517440 | ···×××× | 550200 | ·××···· | 577500 | ·××·×·· | 734580 |
| ·×·×××× | 518980 | ····××× | 550550 | ···×·×× | 579480 | ·×·××·· | 748440 |
| ··××××· | 520520 | ···×·×· | 551320 | ·×·×··· | 582120 | ·×··×·· | 942480 |

The integer $W(X)$ is tabulated against $X$ in the accompanying figure, where we have written $\times$ for $x$ and $\cdot$ for $y$. It is this figure which may be regarded as providing a computer-assisted proof that five weighings suffice for the solution of the problem, since it will be seen that the 63 values of $W(X)$ are all distinct, so that the value of $W(X)$ uniquely determines the set $X$—indeed, it is immediately apparent that the set $X$ is uniquely determined by the first four decimal digits of the six-digit integer $W(X)$. We observe that specification of the third, fourth and fifth digits of $W(X)$ is also sufficient to determine $X$. In order to substantiate this claim we let $WX$ be an APL vector listing the 63 elements of $\{W(X)\colon\ X \in \Sigma\}$ and we evaluate the vector $V$ and the scalar $D$ by the APL command:

$$D \leftarrow \lceil/+/V\circ.=V \leftarrow 10\perp 0\ 0\ 1\ 1\ 1\ 0 \not/ (6\rho 10)\top WX$$

Since $D = 1$ there are no repetitions in the list $V$ of the 8 two-digit and 55 three-digit integers obtained by lopping off the last and the first two digits of the 63 values of $W(X)$.

In practice, of course, the value of $W$ actually observed should be rounded to the nearest integer, or to the nearest element of the set $\{W(X)\colon\ X \in \Sigma\}$. It is this possibility which makes the method robust.

Solved also by H. S. Morse and C. S. Weaver. Partial solutions were received from A. J. Bruni (student), O. P. Lossers (The Netherlands), and M. Pachter (S. Africa).

**Superqueens**

E 3162 [1986, 566]. *Proposed by Paul Monsky, Brandeis University.*

A superqueen is a piece that moves on a square board like an ordinary chess queen but is permitted to continue along the extended diagonals (one may think of the board as a torus with opposite sides next to each other.) A result of Pólya's that has been rediscovered by others from time to time (see, for example, E 2968 [1979, 309]) is that $N$ superqueens may be placed on an $N$ by $N$ board with no two attacking one another if and only if $N$ is prime to 6.

(a) Is it possible, for each value of $N$, to place $N-2$ superqueens on an $N$ by $N$ board with no two attacking one another?

(b) For what values of $N$ can $N-1$ superqueens be so positioned on an $N$ by $N$ board?

*Solution by the proposer.* (a) Such placements are always possible; for each value of $N$ we describe one explicitly. If the $i$th superqueen is in column $x_i$ and row $y_i$ we shall write $q_i$ for the vector $(x_i, y_i)$ viewed as an element of $Z/N \times Z/N$. It suffices to specify the $q_{i+1}-q_i$ for $1 \leqslant i \leqslant N-3$. The "generic" $q_{i+1}-q_i$ will be $(1,2)$ for $N$ odd and $(1,3)$ for $N$ even; however, there will be a few exceptional values of $i$ for which $q_{i+1}-q_i$ takes another value. For conciseness, we denote the vectors $(1,2)$, $(1,3)$, $(1,4)$, $(1,5)$, $(1,6)$, $(2,3)$, $(2,4)$, $(2,5)$, and $(2,1)$ by $A, B, C, D, E, F, G, H$ and $I$, respectively. A notation such as $ABBC$ (or more briefly $AB^2C$) will mean a placement of five superqueens with $q_{i+1}-q_i = A, B, B, C$ for $i = 1,2,3,4$, respectively.

If $N$ is not a multiple of 3 or 4, the solution is easy to describe and in fact more than $N-2$ superqueens are possible (cases 1 and 2 below). The placement of the $N-2$ superqueens is far more elaborate in the remaining cases. The constructions are as follows.

(1) If $N$ is prime to 6, then $A^{N-1}$ places $N$ pairwise non-attacking superqueens.

(2) If $N \equiv \pm 2 \bmod 12$, let $N = 2M$ with $M > 1$. Then $B^{M-1}EB^{M-2}$ places $N-1$ pairwise non-attacking superqueens.

(3) If $N \equiv \pm 3 \bmod 12$, let $N = 6M+3$ with $M > 0$. Then $A^{M-1}BA^{2M}FA^{M-1}FA^{2M-1}$ suffices.

(4) If $N = 12M+4$, then $A$ suffices for $M = 0$ and $B^{2M-1}DB^{3M-1}GB^{M}CB^{3M-1}DB^{3M}$ suffices for $M > 0$.

(5) If $N = 12M+6$, then $BAB$ suffices for $M = 0$ and $B^{2M-1}CB^{4M+1}AFB^{2M-1}HB^{4M}$ suffices for $M > 0$.

(6) If $N = 12M+8$, then $B^{2M}GB^{3M}DB^{M}AB^{3M}GB^{3M+1}$ suffices.

(7) If $N = 12M$, then $A^3BIBA^3$ suffices for $M = 1$ and $B^{3M-2}EB^{M-2}HB^{2M-1}AB^{3M-1}FB^{M-1}CB^{2M-1}$ suffices for $M > 1$.

Setting $z_i = x_i + y_i$ and $t_i = x_i - y_i$, it is straightforward but tedious to show that in each placement above the $x_i$, $y_i$, $z_i$, $t_i$ each represent distinct residue classes mod $N$. Thus no pair of placed superqueens attack each other.

(b) Placing $N-1$ superqueens is possible if and only if $N$ is not divisible by 3 or 4. (1) and (2) of part (a) show the condition is sufficient. To prove necessity,

suppose $N-1$ pairwise nonattacking superqueens have been placed. By translation, we may assume that row $N$ and column $N$ are empty.

Suppose the $i$th superqueen is in row $x_i$ and column $y_i$, and set $z_i = x_i + y_i$ and $t_i = x_i - y_i$. Viewing the summations in terms of $x_i$ and $y_i$, we have $\Sigma z_i = N(N-1)$, $\Sigma t_i = 0$, and $\Sigma(z_i^2 + t_i^2) = 2N(N-1)(2N-1)/3$. Since the superqueens are nonattacking, $\{z_i\}$ and $\{t_i\}$ each occupy $N-1$ congruence classes mod $N$; let $u, v$ with $0 \leqslant u, v < N$ be the omitted class for $\{z_i\}, \{t_i\}$, respectively. Then $u + \Sigma z_i \equiv v + \Sigma t_i \equiv \Sigma_{i=0}^{N-1} i \equiv N(N-1)/2 \bmod N$. Therefore, $u \equiv v \equiv N(N-1)/2 \bmod N$. This implies $u = v = 0$ for odd $N$ and $u = v = N/2$ for even $N$.

When $N$ is odd, we conclude that the $z_i$'s belong to distinct nonzero congruence classes mod $N$; similarly for $\{t_i\}$. Thus $\Sigma z_i^2 \equiv \Sigma t_i^2 \equiv \Sigma j^2 \bmod N$, so $\Sigma(z_i^2 + t_i^2) \equiv N(N-1)(2N-1)/3 \bmod N$. Since the sum equals $2N(N-1)(2N-1)/3$, we conclude that $(N-1)(2N-1)/3$ must be an integer, which implies that $N$ is not divisible by 3 (or 4, being odd).

When $N$ is even, let $z_0 = t_0 = N/2$. Then the $z_i$'s belong to distinct congruence classes mod $N$, and similarly for the $t_i$'s. Since $N$ is even, $z_i \equiv j \bmod N$ implies $z_i^2 \equiv j^2 \bmod 2N$. Summing, we find modulo $2N$ that $\Sigma z_i^2 \equiv \Sigma t_i^2 \equiv \Sigma j^2 = N(N-1)(2N-1)/6$. Deleting the terms for $z_0 = t_0 = N/2$, we have $\Sigma_{i=1}^{N-1}(z_i^2 + t_i^2) \equiv N(N-1)(2N-1)/3 - N^2/2 \bmod 2N$. Again the sum equals $2N(N-1)(2N-1)/3$, and we conclude that $N^2/3 - N/4 + 1/6$ is an integer. Again this implies that $N$ cannot be a multiple of 3 or 4.

No other solutions were received.

### A Multiplicity of Eigenvalues

E 3189 [1987, 181]. *Proposed by M. Lutzky, Silver Spring, MD.*

Prove that the product of two skew-symmetric matrices of order $2N$ has no simple eigenvalues.

*Solution I by O. P. Lossers, Eindhoven University of Technology, Eindhoven, The Netherlands.* Let $A$ and $B$ be skew-symmetric matrices of order $2N$, and let $\lambda$ be an eigenvalue of $AB$. We shall use the following well-known fact (W. Greub, *Linear Algebra*, Springer, 1967, 1975, 1981, §8.15):

$$\text{The rank of a skew-symmetric matrix is even.} \qquad (*)$$

In particular, any skew-symmetric matrix of even order has an even-dimensional kernel.

First suppose $\lambda = 0$. If $B$ is singular, then (*) implies $\dim(\operatorname{Ker} B) \geqslant 2$; because $\operatorname{Ker}(AB) \supseteq \operatorname{Ker} B$, we must then have $\dim(\operatorname{Ker} AB) \geqslant 2$. If $B$ is not singular, then $A$ must be singular, and we have $\operatorname{Ker} AB = B^{-1}(\operatorname{Ker} A)$, which by (*) has dimension at least 2.

Now suppose $\lambda \neq 0$. Let $U = \operatorname{Ker}(ABA - \lambda A)$. Since $\operatorname{Ker}(AB - \lambda I) \supseteq AU$, it suffices to prove $\dim AU \geqslant 2$. By the dimension theorem for linear mappings, we have $\dim AU = \dim U - \dim(U \cap \operatorname{Ker} A)$. The two terms on the right are even, because $\operatorname{Ker} A \subseteq U$ and (*) applies to $A$ and $ABA - \lambda A$. Hence it suffices to show $\operatorname{Ker} A \neq U$.

If $\operatorname{Ker} A = U$, then $(AB - \lambda I)Ax = 0$ implies $Ax = 0$, i.e. $\operatorname{Im} A \cap \operatorname{Ker}(AB - \lambda I) = \{0\}$. On the other hand, $\operatorname{Ker}(AB - \lambda I) \subseteq \operatorname{Im} A$, since $(AB - \lambda I)x = 0$ implies $x = A(\lambda^{-1}Bx)$. This yields a contradiction, because the fact that $\lambda$ is an eigenvalue implies $\operatorname{Ker}(AB - \lambda I) \neq \{0\}$.

*Solution II by A. A. Jagers, Universiteit Twente, Enschede, The Netherlands.* Let $A$, $B$ be skew-symmetric matrices of order $2N$, with $\lambda_0$ an eigenvalue of $AB$ with eigenspace $E$. If $A$ is nonsingular, then $E = \operatorname{Ker}(B - \lambda_0 A^{-1})$, which by (*) has even dimension.

Now suppose $A$ is singular. Let $L$ be the linear space of skew-symmetric complex matrices of order $2N$. The determinant function is a polynomial function on $L$; since it does not vanish everywhere on $L$ it cannot vanish on an open neighborhood of $A$. Therefore, there exist nonsingular skew-symmetric matrices arbitrarily close to $A$.

If $\lambda_0$ is a simple eigenvalue, then the polynomial function $f(M, \lambda) = \det(M - \lambda I)$ on $L \times C$ satisfies $f(AB, \lambda_0) = 0$ and $(\partial f/\partial\lambda)(AB, \lambda_0) \neq 0$. By the implicit function theorem, there is a smooth function $\lambda$ defined on a neighborhood of $AB$ in $L$ such that $\lambda(AB) = \lambda_0$ and $f(M, \lambda(M)) = 0$ on that neighborhood. For $M$ sufficiently close to $AB$, we have $(\partial f/\partial\lambda)(M, \lambda(M)) \neq 0$. However, $M$ can be chosen as $A'B$ for some nonsingular skew-symmetric $A'$ close to $A$, contradicting the statement already proved for $A'B$.

*Editorial comment.* While Losser's solution is the more complicated of the two solutions above, it has the advantage that it is valid for any field of characteristic $\neq 2$.

Solved also by M. Sandy, P. A. Walujo (Indonesia), and the proposer.

**A Failure to Generate 1**

E 3194 [1987, 181]. *Proposed by Ira Gessel, Brandeis University, Waltham, MA.*

Let $S$ be the smallest set of rational functions containing $x$ and $y$ and closed under subtraction and taking reciprocals. Show that $1 \notin S$.

*Solution by O. P. Lossers, Eindhoven University of Technology, Eindhoven, The Netherlands.* Let $U$ be the set of rational functions $f$ such that $f(-x, -y) = -f(x, y)$ for all $x$, $y$. Then $U$ is closed under subtraction and taking reciprocals of non-zero elements. Also, $x, y \in U$, so $S \subset U$. Since $1 \notin U$, we conclude $1 \notin S$.

*Editorial comment.* (1) As noted by Mark Meyerson, the zero function, which belongs to $S$ since $x - x = 0$, does not have a reciprocal. Hence $S$ should be defined to be closed under subtraction and taking reciprocals of nonzero elements, as assumed in the solution above. (2) The proposer informs us that the problem was inspired by Problem 1 of D. J. Newman's *A Problem Seminar* (Springer-Verlag, 1982): "Derive the operations $+$, $-$, $\cdot$, and $\div$ from $-$ and reciprocal." Newman's solution assumes that one is allowed to derive $1 - x$ from $x$, and Gessel's problem shows that this is not possible.

Solved also by S. F. Barger, O. Matouš (Czechoslovakia), L. E. Mattics, M. D. Meyerson, Western Maryland College Problems Group, V. C. Williams and D. J. Williams, and the proposer.

**An Infinity of Natural Limits**

E 3205 [1987, 372]. *Proposed by Dean S. Clark, University of Rhode Island.*

Let

$$a_m = m^{1/2} - \sum_{j=0}^{m} (-1)^{\lfloor j^{1/2} \rfloor},$$

where $\lfloor x \rfloor$ denotes the integral part of $x$. What is the set of sequential limit points of the sequence $\{a_m\}_{m=0}^{\infty}$? In particular, what are $\liminf a_m$ and $\limsup a_m$?

*Composite solution by O. P. Lossers, Eindhoven University of Technology, The Netherlands, and University of South Alabama Problem Group.* The limit points are the non-negative integers and $+\infty$. We first develop a more tractable expression for $a_m$. Since $(-1)^{\lfloor\sqrt{j}\rfloor}$ is constant as $j$ varies from $n^2$ to $(n+1)^2-1$, for any $n$, we obtain

$$\sum_{j=0}^{k^2-1} (-1)^{\lfloor\sqrt{j}\rfloor} = (-1)^{k+1}k.$$

Letting $k = k_m = \lfloor\sqrt{m}\rfloor$ so that $m = k^2 + r$ with $0 \leqslant r = r_m \leqslant 2k_m$, we have

$$\sum_{j=0}^{m} (-1)^{\lfloor\sqrt{j}\rfloor} = (-1)^{k+1}k + (-1)^{k}(r+1),$$

or

$$a_m = k\left(1 + r/k^2\right)^{1/2} + (-1)^k k + (-1)^{k+1}(r+1).$$

Expanding the square root by the binomial formula and collecting terms yields

$$a_m = \left(1 + (-1)^k\right)k + (-1)^{k+1}(r+1) + \frac{r}{2k} + O\left(\frac{1}{m^{1/2}}\right).$$

Let $N_e = \{m \in \mathbb{N}\colon k(m) \text{ is even}\}$ and $N_o = \{m \in \mathbb{N}\colon k(m) \text{ is odd}\}$. By taking subsequences, we need only consider convergent subsequences of $\{a_m\}$ having all indices in $N_e$ or all indices in $N_o$; let $A_e$, $A_o$, respectively, be the limit points of such sequences. We have $a_m = r + 1 + r/2k + O(1/m^{1/2})$ for $m \in N_o$ and $a_m = 2k - r - 1 + r/2k + O(1/m^{1/2})$ for $m \in N_e$. Because $0 \leqslant r/2k \leqslant 1$, we have $A_o = \{1, 2, \dots\} \cup \{\infty\}$ (attained by constant $r$) and $A_e = \{0, 1, \dots\} \cup \{\infty\}$ (attained by constant $2k - r$, in which case $r/2k \to 1$). In particular, $\varliminf a_m = 0$ and $\varlimsup a_m = \infty$.

Solved also by N. Felsinger, J. Fitch, W. Janous (Austria), O. Matouš (Czechoslovakia), A. Pedersen (Denmark), and the proposer. Three incorrect solutions were received.

**A Sequence of Circles**

E 3209 [1987, 457]. *Proposed by Bruce A. Reznick, University of Illinois at Urbana-Champaign.*

Let $C_0, C_1, C_2, \ldots$ be the sequence of circles in the Cartesian plane defined as follows:

(i) $C_0$ is the circle $x^2 + y^2 = 1$,

(ii) for $n = 0, 1, 2, \ldots$ the circle $C_{n+1}$ lies in the upper half-plane and is tangent to $C_n$ as well as to both branches of the hyerbola $x^2 - y^2 = 1$.

Let $r_n$ be the radius of $C_n$. Show that $r_n$ is an integer and give a formula for $r_n$.

*Solution by students of the 1987 Mathematical Olympiad Program, U.S. Military Academy, West Point, NY.* Suppose $n \geqslant 1$. By symmetry, the center of $C_n$ lies on the $y$-axis. Letting $(0, a_n)$ denote the center of $C_n$, the equation of $C_n$ is $x^2 + (y - a_n)^2 = r_n^2$. Because $C_n$ is tangent to $C_{n-1}$, we also have

$$a_n - a_{n-1} = r_n + r_{n-1}. \tag{1}$$

The points of tangency between $C_n$ and the hyperbola $x^2 - y^2 = 1$ correspond to a double root of the equation $y^2 + 1 + (y - a_n)^2 = r_n^2$. A double root must occur at a value where the derivative of $2y^2 - 2ya_n + a_n^2 + 1 - r_n^2$ is 0, i.e., at $y_n = a_n/2$. Substituting for $y$ in the equation, we find

$$a_n^2 = 2r_n^2 - 2. \tag{2}$$

From successive instances of (2), we obtain $a_n^2 - a_{n-1}^2 = 2(r_n^2 - r_{n-1}^2)$. Canceling (1) from this and then adding (1) to what remains yields $2a_n = 3r_n - r_{n-1}$. Substituting this relation into (1) yields the recurrence relation

$$r_n = 6r_{n-1} - r_{n-2}. \tag{3}$$

Since we are given $a_0 = 0$ and $r_0 = 1$, (1) and (2) imply $r_1 = 3$, so that by (3) each $r_n$ is an integer. Solving (3) with $r_0 = 1$, $r_1 = 3$ yields the explicit formula

$$r_n = \frac{1}{2}\left[(3 + 2\sqrt{2})^n + (3 - 2\sqrt{2})^n\right].$$

Solved by more than 50 other readers and the proposer.

### A Positive Function Seeks Its Maximum

E 3226 [1987, 786]. *Proposed by Stan Wagon, Smith College, Northampton, MA.*

The standard derivation of the Wallis product for $\pi$ uses the fact (usually proved by integration by parts) that

$$\lim_{n\to\infty} \frac{\int_0^{\pi/2} \sin^{n+1} x\, dx}{\int_0^{\pi/2} \sin^n x\, dx} = 1.$$

Show that for any continuous function $f$: $[a, b] \to [0, \infty)$ with positive maximum $M$ we have

$$\lim_{n\to\infty} \frac{\int_a^b f(x)^{n+1}\, dx}{\int_a^b f(x)^n\, dx} = M.$$

*Solution by Douglas B. Tyler, University of California, Davis, CA.* Clearly $\overline{\lim}_{n\to\infty}(\int f^{n+1}/\int f^n) \leqslant \overline{\lim}_{n\to\infty}(\int f^n M/\int f^n) = M$. Fix $\varepsilon > 0$, let $E = f^{-1}([M - 2\varepsilon, M])$ and $D = f^{-1}([M - \varepsilon, M])$, so that $E \supseteq D$ and $m(D) > 0$, where $m(D)$ is the Lebesgue measure of the set $D$. Now

$$\frac{(M-2\varepsilon)^n}{\int_E f^n} \leqslant \frac{(M-2\varepsilon)^n}{\int_D f^n} \leqslant \frac{[(M-2\varepsilon)/(M-\varepsilon)]^n}{m(D)},$$

which tends to 0 as $n \to \infty$. Thus

$$\frac{\int_a^b f^{n+1}}{\int_a^b f^n} \geqslant \frac{\int_E f^{n+1}}{\int_E f^n + \int_{\bar{E}} f^n} \geqslant \frac{(M-2\varepsilon)\int_E f^n}{\int_E f^n + m(\bar{E})(M-2\varepsilon)^n}.$$

Dividing the top and bottom by $\int_E f^n$ and letting $n \to \infty$ yields $\underline{\lim}_{n\to\infty}(\int f^{n+1}/\int f^n) \geqslant M - 2\varepsilon$. Since $\varepsilon$ was arbitrary, $\lim_{n\to\infty}(\int f^{n+1}/\int f^n) = M$.

*Editorial comment.* Other solutions or extensions appear in various texts. Several solvers found the problem in Walter Rudin, *Real and Complex Analysis* (McGraw-Hill, 2nd ed., 1974), p. 77 (ex. 23). K.-W. Lau found it in G. Pólya and G. Szegö, *Problems and Theorems in Analysis* (Springer-Verlag), Problem 199 of part II, vol. 1. I. M. Roussos found it in G. DeBarra, *Introduction to Measure Theory* (Van Nostrand Reinhold, 1974), p. 131 (problem 22), with an inaccurate hint on p. 254. E. S. Key showed that if, in the conclusion, $n + 1$ is replaced with $n + c$, then $M$ may be replaced with $M^c$ for any real $c$; for positive $c$ this follows by replacing $f(x)$ by $f(x)^c$ in the assertion of the problem.

Solved by more than 40 other readers and the proposer. Several incorrect or incomplete solutions were received.

### Largest Product for Fixed Sum

E 3230. *Proposed by William Miller, Le Moyne College, Syracuse, New York.*

If $n$ is a given positive integer, what is the largest possible product of distinct positive integers whose sum is $n$?

*Solution by Kenneth Schilling, University of Michigan, Flint.* For all positive integers $i > 2$, let $T_i = 2 + 3 + \cdots + i$. Given a positive integer $n > 4$, find the unique $i$ such that $T_{i-1} \leqslant n < T_i$ and write $n = T_i - j$, where $j \in \{1, 2, \ldots, i\}$. For $n \leqslant 4$, set $S_n = \{n\}$. For $n > 4$, let

$$S_n = \begin{cases} \{3,4,\ldots,i+1\} - \{i\} & \text{if } j = 1; \\ \{2,3,\ldots,i\} - \{j\} & \text{if } j = 2,3,\ldots,i. \end{cases}$$

Then $S_n$ is the set of distinct positive integers with the largest possible product whose sum is $n$.

To prove this, call a set $S$ of positive integers *good* if $S = S_n$ for some $n$. Write $\Sigma S$ and $\Pi S$ as usual for the sum and product, respectively, of the elements of $S$. By construction, for all $n$ there exists exactly one good set $S$ with $\Sigma S = n$. We

complete the proof by constructing, for all sets $S$ which are not good, a set $S'$ of distinct positive integers such that $\Sigma S' = \Sigma S$ and $\Pi S' > \Pi S$.

This is trivial for $\Sigma S \leqslant 4$, so assume from here on that $\Sigma S \geqslant 5$ and that $S = \{s_1, s_2, \ldots, s_m\}$, where $s_1 < s_2 < \cdots < s_m$. Say that $s$ is a *hole* in $S$ if $s_1 < s < s_m$ but $s \notin S$. An alternative characterization of good sets $S$ is

(1) $s_1 = 2$ or 3,
(2) $S$ has at most one hole,
and (3) if $s_1 = 3$ and $S$ has a hole, then that hole is $s_m - 1$.

For each $S$ which fails to satisfy one of the conditions (1)–(3), we shall now construct a set $S'$ such that $\Sigma S' = \Sigma S$ and $\Pi S' > \Pi S$.

Indeed, suppose $S$ does not satisfy (1). If $s_1 = 4$ and $s_2 = 5$, then let $S' = S - \{5\} \cup \{2, 3\}$. If $s_1 = 4$ and $s_2 > 5$, then let $S' = S - \{4, s_2\} \cup \{2, 3, s_2 - 1\}$. If $s_1 > 4$, then let $S' = S - \{s_1\} \cup \{2, s_1 - 2\}$.

Suppose that $S$ does not satisfy (2). Then $S$ has holes $a$ and $b$ such that $a < b$, $a - 1 \in S$, and $b + 1 \in S$. Let $S' = S - \{a - 1, b + 1\} \cup \{a, b\}$. Then $\Pi S' > \Pi S$ because $ab > ab + a - b - 1 = (a - 1)(b + 1)$.

Suppose, finally, that $S$ does not satisfy (3). Then $s_1 = 3$ and $S$ has a hole $a \neq s_m - 1$. By (2), we may assume that $a$ is the only hole in $S$, so in particular $a + 2 \in S$. Let $S' = S - \{a + 2\} \cup \{a, 2\}$. Since $a > 2, 2a > a + 2$, and so $\Pi S' > \Pi S$. This completes the proof.

*Editorial comment.* The related problem obtained by removing the word "distinct" from the statement has made many appearances, including the *Pi Mu Epsilon Journal*, 3 (1959–1964) 119 and 232–233, the 1976 International Mathematical Olympiad, and the 1979 Putnam Competition.

Solved also by N. Artemiadis (Greece), J. C. Binz (Switzerland), D. Callan, J. Delany, L. Erdös (student, Hungary), C. V. Heuer, A. Gorfin, J. Guillerme (France), C. K. Li and B. S. Tam, O. P. Lossers (The Netherlands), R. Persky, D. B. Shapiro, J. H. Steelman, P. Tracy, J. T. Ward, C. H. Webster, and the proposer. Five incorrect or incomplete solutions were received.

## ADVANCED PROBLEMS

6595. *Proposed by P. A. Griffith and J. J. Rotman, University of Illinois at Urbana-Champaign.*

An $R$-module $E$ is *injective* if it is a direct summand of any module containing it. Every module can be imbedded in an injective module. The *injective envelope* $E(M)$ is an injective module containing $M$ such that there is no injective module $E'$ with $M \subset E' \subset E(M)$ and $E' \neq E(M)$. The injective envelope $E(M)$ exists and is unique to isomorphism.

(i) A functor $T$ is called *additive* if, for every pair of $R$-modules $A$ and $B$, we have $T(f + g) = T(f) + T(g)$ whenever $f, g \in \operatorname{Hom}_R(A, B)$. Show that there is a ring $R$ for which there is no additive functor $T$: $R$-mod $\to$ $R$-mod with $T(M) = E(M)$.

(ii) Show that there is a ring $R$ for which there is no functor $T$: $R$-mod $\to$ $R$-mod with $T(M) = E(M)$.

*(iii) Find all rings $R$ for which there exists a functor with the property mentioned in (ii).

6596. *Proposed by S. D. Chatterji and C. El-Hayek, École Polytechnique Fédérale, Lausanne, Switzerland.*

Prove that the differential equation

$$\frac{dx}{dt} = \frac{1}{t^2 + x^2}$$

has a unique solution defined on $\mathbb{R}$ satisfying the initial condition $x(1) = 1$.

6597. *Proposed by Ichiro Murase, Science University of Tokyo, Japan.*

Let $f(x) \in \mathbb{R}[x]$, say

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n, \qquad n \geqslant 0, \qquad a_n \neq 0.$$

Let $Z_+(f)$ denote the number of zeros of $f$ in $(0, +\infty)$, each counted according to its multiplicity. Let $W(a_0, a_1, \ldots, a_n)$ denote the number of pairs $(a_k, a_l)$ with $k < l$ such that $a_k a_l < 0$ and $a_v = 0$ for $k < v < l$. Descartes' rule of signs says that $Z_+(f) \leqslant W(a_0, a_1, \ldots, a_n)$ and $W(a_0, a_1, \ldots, a_n) - Z_+(f)$ is even.

If there is an index $k$, $1 \leqslant k \leqslant n-1$, such that

$$a_{k-1}a_k < 0,\ a_k a_{k+1} < 0, \qquad a_k^2 \leqslant a_{k-1}a_{k+1},$$

prove that

$$Z_+(f) \leqslant W(a_0, a_1, \ldots, a_n) - 2.$$

## SOLUTIONS OF ADVANCED PROBLEMS

6544 [1987, 387]. *Proposed by Daniel Goffinet, Lycée Claude Fauriel, St. Etienne, France.*

What is the partial fraction decomposition of $1/(x^n - 1)$ over the field $Q$ of rational numbers?

*Solution by O. P. Lossers, Eindhoven University of Technology, The Netherlands.* The factorization of $x^n - 1$ into irreducible monic polynomials is

$$x^n - 1 = \prod_{d \mid n} \phi_d(x),$$

where $\phi_d$ denotes the cyclotomic polynomial, i.e., the monic polynomial of degree $\varphi(d)$ that has the primitive $d$th roots of unity as its zeroes. Logarithmic differentiation yields

$$\frac{nx^{n-1}}{x^n - 1} = \sum_{d \mid n} \frac{\phi_d'(x)}{\phi_d(x)},$$

or

$$n + \frac{n}{x^n - 1} = \sum_{d|n} \frac{x\phi_d'(x)}{\phi_d(x)},$$

from which it follows that

$$\frac{1}{x^n - 1} = \frac{1}{n} \sum_{d|n} \frac{x\phi_d'(x) - \varphi(d)\phi_d(x)}{\phi_d(x)}.$$

Solved also by David G. Cantor, William H. Gustafson, Keith A. Kearnes, Douglas B. Tyler, and the proposer.

6545 [1987, 469]. *Proposed by L. A. Rubel, University of Illinois at Urbana-Champaign.*

Suppose $p_n(z)$ is a polynomial over $\mathbb{C}$ for $n = 0, 1, 2, \ldots$ and suppose $\sum_{n=0}^{\infty} p_n(z) w^n$ converges for all complex $z$ and $w$. Must $\sum_{n=0}^{\infty} p_n'(z) w^n$ converge for all complex $z$ and $w$? (Here $p_n'(z)$ is the ordinary derivative of $p_n(z)$.)

*Solution by the proposer.* No. In fact, the polynomials $p_n(z)$ can be chosen so that

$$g(w) = \sum_{n=0}^{\infty} p_n'(0) w^n$$

does not converge for any nonzero $w$. The idea is to find polynomials that are "small" on a sequence of large open sets that expand to exhaust the left and right half planes, while they are "large" on a sequence of small open sets that contract to the origin. For present purposes define $n!! = (n!)!$; thus, e.g., $5!! = 1 \cdot 2 \cdot \cdots 120$.

Let

$$\begin{aligned}
B_n &= \{z\colon |\mathrm{Im}(z)| \leqslant 4/n!!, \mathrm{Re}(z) \geqslant -4/n!!\},\\
R_n &= \{z\colon 1/n \leqslant \mathrm{Re}(z) \leqslant 2n, -n \leqslant \mathrm{Im}(z) \leqslant n\},\\
L_n^* &= \{z\colon -2n \leqslant \mathrm{Re}(z) \leqslant 1/(2n), -n \leqslant \mathrm{Im}(z) \leqslant n\},\\
L_n &= L_n^* \setminus B_n
\end{aligned}$$

and

$$M_n = \{z\colon |\mathrm{Re}(z)| \leqslant 1/n!!, |\mathrm{Im}(z)| \leqslant 1/n!!\}.$$

Roughly speaking, a small "bite" $B_n$ has been taken out of $L_n^*$, into which has been placed the small set $M_n$ for mischievous purposes. Clearly we may surround $R_n$, $L_n$, and $M_n$ by open sets no two of which intersect, and whose union does not separate the plane. Hence we may apply Runge's Theorem (cf. Saks-Zygmund, 2nd edition, Chapter IV, Theorem 2.2) to choose a sequence of polynomials $r_n(z)$ such that

$$|r_n(z)| \leqslant 1/(3n), \qquad z \in L_n \cup R_n,$$

and

$$|r_n(z) - n!!| < 1, \qquad z \in M_n.$$

Set $q_n(z) = z r_n(z)$. Then $q_n'(0) = r_n(0)$ must exceed $n!! - 1$ in modulus, while

$$|q_n(z)| < 1, \qquad z \in L_n \cup R_n.$$

Set $p_n(z) = q_n(z)/n!$ Then $p_n'(0)$ must exceed $(n!! - 1)/n!$ in modulus, so that $g(w)$ has the claimed property. On the other hand, set

$$f(z, w) = \sum_{n=0}^{\infty} p_n(z) w^n.$$

Since $p_n(0) = q_n(0) = 0$, we may assume $z \neq 0$. Hence, if $\operatorname{Re}(z) \leqslant 0$, then $z \in L_n$ for $n$ sufficiently large. But if $\operatorname{Re}(z) > 0$, then $z \in R_n$ for $n$ sufficiently large. Hence for $z$ non-zero

$$|p_n(z)| < 1/n!$$

if $n$ is sufficiently large; the result follows since the series $f(z, w)$ always converges.

*Editorial Comment.* Leonard Lipshitz of Purdue University has pointed out that it follows from a result of Abhyankar and Moh (see [**1**]) that if $\sum_{n=0}^{\infty} p_n(z) w^n$ converges as a power series in $w$ in some neighborhood of $w = 0$ for each $z \in M$, where $M$ is a set of positive two-dimensional Lebesgue measure in $\mathbb{C} = \mathbb{R}^2$, then the series that results on grouping $\sum p_n(z) w^n$ as a power series in $z$ and $w$ must be convergent in a neighborhood of $(z, w) = (0, 0)$, *provided* that $\deg p_n \leqslant n$ for all $n$.

The proposer's solution shows that, without this additional restriction, the result fails, in general. It is easy to modify the Lipshitz remark to include the case $\deg p_n = O(n)$. (Just replace $w$ by $w^K$ for a suitable $K$.) It would be interesting to see whether the hypothesis $\deg p_n < \varphi(n)$ implies the result for any $\varphi$ for which $\varphi(n) \neq O(n)$, and if so, for which $\varphi$.

The main result of [1], tailored to the context of two complex variables, asserts that if $f(x, y) = \sum_{i=0}^{\infty} a_{ij} x^i y^j$ is a formal power series in the two complex variables $x$ and $y$, and if, for all $s$ in a subset $S$ of $\mathbb{C}$ of positive measure, $f(sy, y)$ has a positive radius of convergence as a power series in $y$, then $f(x, y)$ converges in some neighborhood of the origin in $\mathbb{C}^2$.

In more specific form Lipshitz's remark is the following.

THEOREM. *Let $f(u, v) = \sum a_{ij} u^i v^j$ where $a_{ij} = 0$ if $i > j$. Suppose that for every $\bar{u} \in \mathbb{C}$ (or a subset $M$ of $\mathbb{C}$ of positive measure) $f(\bar{u}, v)$ converges as a power series in $v$. Then $f(u, v)$ converges as a power series in two variables.*

*Proof.* Let $u = x/v$. Then

$$f\left(\frac{x}{v}, v\right) = \sum a_{ij} \left(\frac{x}{v}\right)^i v^j = \sum a_{ij} x^i v^{j-i}.$$

Since $a_{ij} = 0$ for $i > j$ this is a power series in $x$ and $v$. Call it $g(x, v)$. If we set $x = \alpha v$, where $\alpha \in M$, then

$$g(\alpha v, v) = \sum a_{ij} (\alpha v)^i v^{j-i} = \sum a_{ij} \alpha^i v^j = f(\alpha, v).$$

Hence $g(\alpha v, v)$ is a convergent power series in $v$ for all $\alpha$ in the set $M$ of positive measure. Hence by Abhyankar and Moh $g(x, v)$ is convergent as a power series in

two variables. So $\Sigma a_{ij}x^i v^{j-i} = g(x, v)$ is convergent. Hence so is

$$g(uv, v) = \sum a_{ij}(uv)^i v^{j-i} = \sum a_{ij}u^i v^j = f(u, v).$$

This completes the proof.

REFERENCE

1. S. S. Abhyankar and T. T. Moh, A reduction theorem for divergent power series, *J. reine angew. Math.*, 241 (1970) 27–33.

The proposer's solution was the only correct solution received.

# REVIEWS

Edited by Joseph Konhauser
*Department of Mathematics, Macalester College, St. Paul, MN 55105*

*Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds.* By G. Pólya and R. C. Read. Springer-Verlag, New York, 1987, vi + 148 pp.

Russell Merris[1]
*Mathematics and Computer Science Department, California State University, Hayward, California 94542*

Suppose you are given three colors, say red, white, and blue, with which to paint the faces of the die in Figure 1. The only rule is that each face must be uniformly painted one of the 3 colors. In how many different ways can the job be done? The answer depends on how we interpret the word "different." Imagine two possibilities. In the first case, the numbers on the die faces show through the paint, so that if face 1 is painted blue and the other 5 faces are painted red, we can tell this is different from painting face 4 blue and all the others red. Here the answer is an easy application of the so-called Fundamental Counting Principle, namely, $3^6 = 729$.

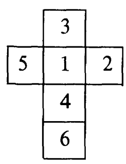



Fig. 1

The more interesting case occurs when the numbers disappear under the paint. Once the dice have been rolled around a few times, it makes very little difference whether it was face 1 or face 4 that was originally distinguished by the blue coat. We will call two (of the 729) colorings *equivalent* if one can be rotated so as to appear identical to the other. An equivalence class of colorings is a color *pattern*. If we interpret "different" as meaning inequivalent, then it is the number of patterns that we seek (a number which would have warmed the heart of H. J. Heinz). Before going on, the reader may wish to convince himself that the colorings illustrated in Figures 2 and 3 are inequivalent.

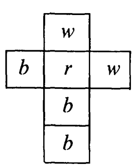



Fig. 2

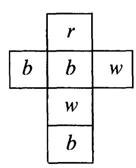



Fig. 3

[1]Supported by ONR contract 85-K-0335.

Formally, we may describe the coloring in Figure 2 by means of a function $f$: $\{1,2,3,4,5,6\} \to \{r, w, b\}$, specifically, $f = (r, w, w, b, b, b)$. Alternatively, one may identify $f$ with the "word" *rwwbbb*, which one is tempted to write $rw^2b^3$. A problem with yielding to this temptation is that the inequivalent coloring $g$ in Figure 3 may be expressed using the word *bwrwbb* which, if written as $rw^2b^3$, might easily be confused with $f$. Pólya's great idea was to welcome this confusion. Think of a word like *bwrwbb* as a product of noncommuting variables. If the variables are allowed to commute, we obtain an expression of the form $rw^2b^3$ which we might call the *weight* of the word. Clearly, equivalent colorings have the same weight.

We can, of course, count the number of 6-letter words of weight $rw^2b^3$. It is just the multinomial coefficient

$$\frac{6!}{1!2!3!} = 60.$$

So, of the 729 colorings, 60 have weight $rw^2b^3$. Indeed, the 60 colorings of this weight comprise a union of (at least 2) color patterns. Given 3 nonnegative integers $i$, $j$ and $k$ that add up to 6, let $c_{i,j,k}$ be the number of *inequivalent* red-white-blue colorings of the cube with weight $r^iw^jb^k$. Then, for example, $c_{5,0,1} = 1$ and $c_{1,2,3} \geqslant 2$. Moreover, by symmetry, $c_{1,2,3} = c_{3,2,1} = c_{1,3,2} = \cdots$.

The *pattern enumerator* is the generating function

$$P(r, w, b) = \sum c_{ijk} r^i w^j b^k.$$

Pólya's Theorem amounts to a clever way of obtaining $P(r, w, b)$ from the rotation group $G$ of the cube:

| | | | |
|---|---|---|---|
| identity | (1265) | (12)(34)(56) | (123)(465) |
| | (1364) | (13)(25)(46) | (124)(365) |
| | (1463) | (14)(25)(36) | (132)(456) |
| (16)(25) | (1562) | (15)(26)(34) | (135)(264) |
| (16)(34) | (2354) | (16)(23)(45) | (142)(356) |
| (25)(34) | (2453) | (16)(24)(35) | (145)(263) |
| | | | (153)(246) |
| | | | (154)(236) |

Observe that a function (coloring, word) $h$ is equivalent to $f$ if and only if $h = f\sigma$ for some rotation $\sigma \in G$. For purposes of enumeration what turns out to be important is the cycle structure of $\sigma$. If $c_i(\sigma)$ is the number of cycles of length $i$ in the disjoint cycle factorization of $\sigma$, define

$$m(\sigma) = \prod_{i=1}^{6} s_i^{c_i(\sigma)}.$$

Then $m(\sigma)$ is a monomial in the 6 variables $s_1, s_2, \ldots, s_6$; it is a device for keeping track of the cycle structure of $\sigma$. For example, $m(2453) = s_1^2 s_4$ (don't forget the cycles of length 1), and $m(16)(34) = s_1^2 s_2^2$. The arithmetic mean of these 24 monomi-

als is the *cycle index polynomial* of $G$,

$$
\begin{aligned}
Z_G(s_1, s_2, \dots, s_6) &= \frac{1}{24} \sum_{\sigma \in G} m(\sigma) \\
&= \frac{1}{24} \sum_{\sigma \in G} \prod_{i=1}^{6} s_i^{c_i(\sigma)} \\
&= \frac{1}{24}\left(s_1^6 + 3s_1^2 s_2^2 + 6s_1^2 s_4 + 6s_2^3 + 8s_3^2\right).
\end{aligned}
$$

Note that, in our case, $Z_G$ actually involves only $s_1$, $s_2$, $s_3$ and $s_4$.

We are now ready to state Pólya's Counting Theorem: The pattern enumerator is obtained from the cycle index polynomial by substituting for $s_i$ the $i$th power sum of $r$, $w$ and $b$, i.e.,

$$s_i = r^i + w^i + b^i.$$

In our particular case,

$$
\begin{aligned}
&P(r, w, b) \\
&\quad = \frac{1}{24}\Big[(r+w+b)^6 + 3(r+w+b)^2(r^2+w^2+b^2)^2 \\
&\qquad + 6(r+w+b)^2(r^4+w^4+b^4) + 6(r^2+w^2+b^2)^3 + 8(r^3+w^3+b^3)^2\Big] \\
&\quad = \{6\} + 1\{5,1\} + 2\{4,2\} + 2\{4,1,1\} + 2\{3,3\} + 3\{3,2,1\} + 6\{2,2,2\},
\end{aligned}
$$

where

$$
\begin{aligned}
\{6\} &= r^6 + w^6 + b^6, \\
\{5,1\} &= r^5 w + r^5 b + r w^5 + r b^5 + w^5 b + w b^5, \\
\{4,2\} &= r^4 w^2 + r^4 b^2 + r^2 w^4 + r^2 b^4 + w^4 b^2 + w^2 b^4, \\
\{4,1,1\} &= r^4 w b + r w^4 b + r w b^4, \\
\{3,3\} &= r^3 w^3 + r^3 b^3 + w^3 b^3, \\
\{3,2,1\} &= r^3 w^2 b + r^3 w b^2 + r^2 w^3 b + r^2 w b^3 + r w^3 b^2 + r w^2 b^3,
\end{aligned}
$$

and

$$\{2,2,2\} = r^2 w^2 b^2.$$

Note, e.g., that the coefficient of $rw^2b^3$ in $P(r, w, b)$ is 3. This means there are exactly 3 color patterns of this weight. Figures 2 and 3 illustrate representatives from 2 of them. By the same token, there must be 6 color patterns of weight $r^2w^2b^2$.

Since we now know the number of color patterns of each weight, it is a simple matter to compute the total number of color patterns. Observe that this grand total

may be expressed as

$$P(1,1,1) = \frac{1}{24} \sum_{\sigma \in G} \prod_{i=1}^{6} (1^i + 1^i + 1^i)^{c_i(\sigma)}$$

$$= \frac{1}{24} \sum_{\sigma \in G} 3^{c(\sigma)}$$

$$= \frac{1}{24} \left[ 3^6 + 3(3^2)(3^2) + 6(3^2)(3) + 6(3^3) + 8(3^2) \right] = 57,$$

where $c(\sigma) = c_1(\sigma) + \cdots + c_6(\sigma)$ is the total number of cycles in $\sigma$.

Published in German in 1937, Pólya's now classic paper contained one theorem and 100 pages of applications [**4**]. As the catalyst for a whole new theory of combinatorial enumeration, Pólya's article is one of the most significant papers in 20th-century mathematics. Now, on the 50th anniversary of its appearance, the article has been republished in English translation with an outstanding supplement by Ronald Read.

It turns out that Pólya's theorem (1937) was first published by J. H. Redfield (1927), but no one understood Redfield's paper until it was explained by F. Harary (1960). Conceptually, the theorem is a version of what has come to be known as Burnside's Lemma, which is due to Cauchy but attributed by Burnside to Frobenius! In addition to historical anecdotes, Read shows us the facade of the mathematical edifice which has been built on the foundation of Pólya's work, and provides us with approximately 250 references. Among these, the single one most worthy of mention was written by N. G. de Bruijn [**2**]. In the absence of an English translation of the original, many of us learned Pólya's theorem from de Bruijn's lucid exposition. In addition, [**2**] contains a significant extension which incorporates a second group of symmetries operating on the "colors." A more recent extension involves the introduction of group characters into the cycle index polynomial with applications to the NMR spectrum [**1**; **6**]. Among the many approaches to Pólya's Theorem, one of the least expected involves Möbius inversion [**5**]. Finally, an article of special interest to the reviewer involves the distance between color patterns [**3**].

REFERENCES

1. K. Balasubramanian, Applications of combinatorics and graph theory to spectroscopy and quantum chemistry, *Chem. Rev.*, 85 (1985) 599–618.
2. N. G. de Bruijn, Pólya's theory of counting, in Applied Combinatorial Analysis (E. F. Beckenbach, ed.), Wiley, NY, 1964.
3. R. Merris, Generalized matrix functions and pattern inventory, *Linear & Multilinear Algebra*, 12 (1983) 315–327.
4. G. Pólya, Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen, *Acta Mathematica*, 68 (1937) 145–254.
5. G.-C. Rota and D. A. Smith, Enumeration under group action, *Ann. Scuola Normale Superiore-Pisa*, 4 (1977) 637–646.
6. S. G. Williamson, Symmetry operators of Kranz products, *J. Combinatorial Th.* 11 (1971) 122–138.

*From One to Zero: A Universal History of Numbers.* By Georges Ifrah (translated by Lowell Bair), Viking Penguin Inc., 1987. xvi + 503 pp.

FRANK SWETZ
*Departments of Mathematics and Education, Pennsylvania State University, Harrisburg, PA 17057*

I have just read a beautiful book. This is a term that I use to describe a painting or a poem but seldom a book, especially one on mathematics. However, in this instance, I feel it is justified. Georges Ifrah's book, *From One to Zero*, presents a history of numbers or, more correctly, numerals, how they were conceived, transcribed and how they eventually evolved into the system we know today as the Hindu-Arabic numerals. In the past, other authors have undertaken the same feat and produced noteworthy works. Tobias Dantzig's *Number: The Language of Science* (1930) and Karl Menninger's *Number Words and Number Symbols* (U.S. edition, 1969) come easily to mind. Ifrah's most recent contribution to this literature is truly engrossing. He supplements and extends the work of his predecessors, updates research findings, fills in missing information and develops relevant theories that cast historical and sociological insights into the place of numeration and the use of numerals in our cultural heritage.

Prompted by a child's questions, "Where do figures come from?," "When did people learn to count?," and "How did numbers start?," Georges Ifrah began his quest to learn the answers to these questions. The result is a systematic explanation and documentation of the evolution of our numeral system. The book has six major divisions: Awareness of Numbers; Concrete Counting; Invention of Numerals, Numerals and Letters, Hybrid Numeration Systems and The Ultimate State of Numerical Notation, which span twenty-nine chapters that focus on specific aspects of numeric development. Ifrah is a French citizen, fluent in both Arabic and Hebrew, a teacher by profession and a graphic artist by avocation. All these attributes contributed to the production of this book. In particular, Ifrah's quest for information utilized findings available in French research literature, an extremely rich resource of anthropological discoveries and comments on the cultural ascendancy of man and one that is frequently overlooked by English-speaking researchers. His abilities in Arabic and Hebrew allowed for an in-depth examination of numerical usage and tendencies in these two societies. For example, in an examination of magic squares several years ago, I came across a medieval Cabalistic rendering of the sacred name Yahweh embedded in a magic square of order three but could not determine how the letters of the Hebrew alphabet were assigned their numerical values. I found the answer in Ifrah's chapter 21 on the consideration of "Numerals, Letters, Magic, Mysticism, and Divination."

In reading this book, many such answers were found and new levels of understanding reached. Numerals are visual expressions of a number sense and as such they must be examined and analyzed visually as well as theoretically. *From One to Zero* abounds with beautifully rendered illustrations of which 395 comprise formal figures while numerous others are incorporated in, and referred to, throughout the text. Ifrah's skill as an artist serves the reader well—the evolution of tally notches to symbolic Roman numerals is graphically demonstrated as is the graphic progression of Egyptian hieroglyphic numerals to their cursive hieratic form. Such demonstra-

tions are interspersed throughout the historical discussion providing a visual dimension of understanding into the evolutionary process of symbolic number rendering.

In its entirety, the book stands as a testimony to Georges Ifrah the teacher. The reader is drawn into the developments described. Upon reading Chapter 3, "The First Calculating Machine: The Hand," I found myself practicing an Afghan duodecimal finger-counting technique and performing digital contortions in isolating the Chinese finger position for 100,000. An intriguing premise is put forward in the question title of chapter 10, "Was Writing Invented by Accountants?," in which the ancient Sumerian practice of using concrete token impressions on wet clay for numerical accounting is examined as a precursor of writing. If this premise is carried forward, mathematics is the mother not only of science but also of literature!

*From One to Zero*: *A Universal History of Numbers* was originally published in French under the title *Histoire Universelle des Chiffres* (1981). Lowell Bair's English-language translation was produced by Viking Penguin and released in hard cover in 1985. The 1987 edition is a paperback version of the work in its entirety. This book is an outstanding piece of scholarship, meticulously conceived and crafted. It is beautifully designed and highly readable. Unfortunately, it lacks an index. The book is highly recommended personal reading for anyone curious about the origins of counting, number uses and symbols, and the beginnings of mathematics itself. A prize reference for university and high school libraries.

# TELEGRAPHIC REVIEWS

*Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.*

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

| | | |
|---|---|---|
| T: Textbook | P: Professional Reading | 1–4: Semesters |
| C: Computer Software | L: Undergraduate Library | **: Special Emphasis |
| S: Supplementary Reading | 13: Grade Level | ??: Questionable |

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

**General, P, L*.** *Izrail M. Gelfand: Collected Papers, Volume II.* Ed: S.G. Gindikin, *et al.* Springer-Verlag, 1988, x + 1039 pp, $149.50. [ISBN: 0-387-13619-3] Second of three volumes (*Volume I,* TR, December 1988) featuring 75 papers on group representation theory arranged in nine thematic groups. All papers appear in English. Contains a complete Gelfand bibliography, keyed to the three volumes. LAS

**General, S(13), P, L.** *The Puzzling Adventures of Dr. Ecco.* Dennis Shasha. WH Freeman, 1988, xiii + 181 pp, $9.95 (P); $15.95. [ISBN: 0-7167-1978-9; 0-7167-1958-4] A book about puzzles and problem solving, disguised as the fictional tale of Dr. Jacob Ecco, the legendary private detective. Readers are invited to solve thirty-eight puzzles posed in narrative form. An effective device for honing thinking and problem solving skills. Also entertaining. CEC

**General, P.** *The Mathematical Heritage of Hermann Weyl.* Ed: R.O. Wells, Jr. Proc. of Symp. in Pure Math., V. 48. AMS, 1988, vii + 344 pp, $47. [ISBN:0-8218-1482-6] Eighteen papers honoring Weyl's contributions to mathematics and mathematical physics, especially group representation theory. An impressive list of authors—many may be the object of future "heritage" volumes. BC

**Mathematics Appreciation, S*, L*.** *Penrose Tiles to Trapdoor Ciphers.* Martin Gardner. WH Freeman, 1989, ix + 311 pp, $13.95 (P); $19.95. [ISBN: 0-7167-1987-8; 0-7167-1986-X] Fractals, surreal numbers, Ramsey theory, paradoxes, and much more from the inexhaustible spring of Gardner's *Scientific American* column. A few new chapters extend original columns (on Penrose tiles, the Oulipo, and trapdoor codes), and a concluding chapter re-engages Dr. Matrix to pose some amazing facts about magic squares. LAS

**Education, P*, L**.** *Proceedings of the Sixth International Congress on Mathematical Education.* Ed: Ann and Keith Hirst. ICMI Secretariat. János Bolyai Mathematical Society, (U.S. Distr: Mathematical Association of America), 1988, 397 pp, $45. [ISBN: 963-8022-48-5] Five plenary addresses (B. Nebres on school mathematics in developing countries; G. Vergnaud on psychology of mathematics education; A. Ershov on computerization of mathematics education in the U.S.S.R.; L. Lovász on algorithmic mathematics, and J.-P. Kahane on George Pólya) followed by careful reports from each of the 30 action, theme, and topic groups that met at ICME-VI in Budapest in August, 1988. A wealth of up-to-date information about mathematics education around the world, published in record-breaking time of four months. All but Kahane's paper are in English. LAS

**Education, P, L.** *Managing Mathematics: A Handbook for the Head of Department.* The Mathematical Association and Stanley Thornes, 1988, xii + 146 pp, £4.25 (P). [ISBN: 0-85950-918-4] Written as a consequence of recommendations in the Cockcroft report *Mathematics Counts,* this slim volume presents sound advice and reminders of things to consider for secondary school department heads. Although many details are unique to Great Britain, much of the wisdom of this handbook applies more broadly. Well written; comprehensive; useful. LAS

**Education, P*, L**.** *Everybody Counts: A Report to the Nation on the Future of Mathematics Education.* National Research Council. National Academy Press, 1989, xii + 114 pp, $7.95. [ISBN: 0-309-03977-0] A forceful consensus statement from three mathematics units at the National Academy of Sciences (MSEB, BMS, MS2000) on mathematics education "as all one system, from kindergarten through graduate school." Ranging from human resources and national needs to curricula and teaching, the report outlines a national strategy for responding to the many challenges facing mathematics today. According to the Preface, *Everybody Counts* is intended as a signal that the nation's leading scientists and engineers—the "Academy"—are prepared

to join with mathematicians and mathematics educators "to participate actively in the long-term work of rebuilding mathematics education in the United States." LCL

**History, S(13-16), L.** *Descartes' Dream: The World According to Mathematics.* Philip J. Davis, Reuben Hersh. Penguin Books, 1988, xviii + 321 pp, $8.95 (P). [ISBN: 0-14-022787-3] A collection of independent essays centered around several different themes. Among the topics approached are "...the impact mathematics makes when it is applied to the world of nature or of human activities" and "how does the computerization of the world affect the physical and intellectual quality of civilization?" The essays require different levels of mathematical knowledge, ranging from very little to a fairly high level. A very interesting, well-written book. Paperback reprint of 1986 Harcourt Brace hardcover (TR, March 1987). RH

**History, P.** *Mathematics from Manuscript to Print: 1300-1600.* Ed: Cynthia Hay. Clarendon Pr, 1988, ix + 273 pp, $67. [ISBN: 0-19-853909-6] "'Renaissance mathematics? Was there any?'" This volume records the proceedings of a joint British-French conference on mathematics history during the period indicated, recognizing the quincentenary of Chuquet's *Triparty.* Eighteen papers (in English): Italian and Provencal mathematics; Chuquet and French mathematics; sixteenth century mathematics; ramifications of mathematics and its history. RB

**Logic, P.** *Lecture Notes in Mathematics-1292: Classification Theory.* Ed: J.T. Baldwin. Springer-Verlag, 1987, 500 pp, $48.50 (P). [ISBN: 0-387-18674-3] Proceedings of U.S.-Israel Workshop on Model Theory in Mathematical Logic held in Chicago, December 15-19, 1985. Research papers and surveys of recent work. Introduction provides brief history of subject and overview of major directions of research. KS

**Logic, P.** Logic, P. *Lecture Notes in Mathematics-1333: Cabal Seminar 81-85.* Ed: A.S. Kechris, D.A. Martin, J.R. Steel. Springer-Verlag, 1988, 224 pp, $20 (P). [ISBN: 0-387-50020-0]

**Foundations, T(18: 1), P.** *Introduction to Higher Order Categorical Logic.* J. Lambek, P.J. Scott. Stud. in Adv. Math., V. 7. Cambridge U Pr, 1988, ix + 293 pp, $24.95 (P). [ISBN: 0-521-35653-9] Paperback edition (with corrections) of 1986 publication (TR, April 1987). KS

**Discrete Mathematics, T(13-14: 1).** *Introduction to Discrete Mathematics.* Steven C. Althoen, Robert J. Bumcrot. PWS-Kent, 1988, xii + 346 pp, $27.50. [ISBN: 0-534-91504-3] Written for the future computer scientist, the text presents topics unified by the use of algorithms. Chapter one begins by describing algorithms; subsequent chapters present enumeration, induction, recursion, circuits, trees, Eulerian graphs, planarity, and networks. Includes a section on algorithm efficiency and one on error propagation. The final chapter introduces relations, algebraic structures, and machines. Written at a very elementary level; contains many diagrams, examples, and exercises. LW

**Number Theory, P.** *Investigations in Number Theory.* Ed: T. Kubota. Adv. Stud. in Pure Math., V. 13. Academic Pr, 1988, 621 pp, $89.50. [ISBN: 0-12-427640-7] Nineteen papers on class field theory, diophantine inequalities, and automorphic forms. Includes a numerical table of class numbers and fundamental units for pure cubic fields. BC

**Number Theory, P*.** *Number Theory and Its Applications in China.* Ed: Wang Yuan, Yang Chungchun, Pan Chengbiao. Contemp. Math., V. 77. AMS, 1988, xiii + 170 pp, $20 (P). [ISBN: 0-8218-5084-9] This book contains nine survey articles and three articles on recent research in the People's Republic. The research accomplishment of Chinese number theorists from 1949-1979 are emphasized. Wide ranging, with some excellent lists of references. CEC

**Group Theory, P.** *Nineteen Papers on Algebraic Semigroups.* A. Ya. Aĭzenshtat, *et al.* AMS Transl. Ser. 2, V. 139. AMS, 1988, vi + 210 pp, $69. [ISBN: 0-8218-3115-1] Papers on algebraic semigroups by Aĭzenshtat, Podran, Ponizovskiĭ, Shaĭn, Shutov, and Vazhenin. Translated from Russian. RH

**Algebra, P.** *The Structure of Finite Algebras.* David Hobby, Ralph McKenzie. Contemp. Math., V. 76. AMS, 1988, xi + 203 pp, $27 (P). [ISBN: 0-8218-5073-3] A finite algebra is a finite set of elements together with a (possibly infinite) set of operations acting on this set of elements. Finite groups and rings are examples of finite algebras. The main result presented in this book is that the lattice of congruences of a finite algebra determines much of the structure of that algebra. Other new results are also presented. RH

**Calculus, T(13: 2).** *Calculus, Second Edition.* J. Douglas and Barbara T. Faires. Math. Ser. Random House, 1988, xx + 1196 pp, $40. [ISBN: 0-394-36624-7] A standard calculus text which is particularly well done. All the graphs have been redrawn for this edition. Readable, with lots of worked examples, routine exercises, and challenging problems. The two-color format is particularly nice. (*First Edition,* TR, June-July 1983.) CEC

**Complex Analysis, T(18: 1, 2), S, P.** *Capacities in Complex Analysis.* Urban Cegrell. Aspects of Math., E. 14. Friedr. Vieweg & Sohn, 1988, xi + 153 pp, DM 42 (P). [ISBN: 3-528-06335-1] A study of capacity theory for plurisubharmonic and analytic functions of several complex variables. In the well-understood case of one complex variable, the Laplace operator links function theory to the capacity of sets; here, the link is the complex Monge-Ampere operator. In twelve brief chapters, each with references. PZ

**Differential Equations, P.** *Solution of Variational Inequalities in Mechanics.* I. Hlaváček. Appl. Math. Sci., V. 66. Springer-Verlag, 1988, x + 275 pp, $42.80 (P). [ISBN: 0-387-96597-1] Theory and numerical techniques for variational inequalities, with empha-

sis on plasticity and contact problems of elastic bodies. BC

**Partial Differential Equations, P.** *Lecture Notes in Mathematics-1340: Calculus of Variations and Partial Differential Equations.* Ed: S. Hildebrandt, D. Kinderlehrer, M. Miranda. Springer-Verlag, 1988, 301 pp, $28.60 (P). [ISBN: 0-387-50119-3] Proceedings of a June 1986 conference held in Trento, Italy in honor of Hans Lewy. 23 diverse papers, arranged alphabetically by author. LAS

**Numerical Analysis, L.** *Numerical Analysis 1987.* Ed: D.F. Griffiths, G.A. Watson. Pitman Res. Notes in Math. Ser., V. 170. Longman Scientific & Technical (US Distr: Wiley), 1988, 300 pp, $54.95 (P). [ISBN: 0-470-21012-5] Sixteen invited papers presented at the 12th Dundee Biennial Conference on Numerical Analysis held on June 23-26, 1987. The papers cover a broad range of general numerical techniques. SM

**Functional Analysis, P.** *Lecture Notes in Mathematics-1302: Function Spaces and Applications.* Ed: M. Cwikel, *et al.* Springer-Verlag, 1988, vi + 445 pp, $39.40 (P). [ISBN: 0-387-18905-X] Proceedings of a seminar in Lund, Sweden, June 1986. In keeping with M. Riesz's contributions, the topic is mostly interpolation theory and its current impact on function space methods in analysis. Contains thirty-five papers and a problem section. TAV

**Functional Analysis, P.** *Topological Algebras: Selected Topics.* Anastasios Mallios. Math. Stud., V. 124. North-Holland (US Distr: Elsevier Science), 1986, xvii + 535 pp, $64.75 (P). [ISBN: 0-444-87966-8] Almost entitled *General Theory of Topological Algebras*, this dense text is intended for those who wish to apply the methods and results of this theory to other disciplines. The author assumes some familiarity with the general theory of topological vector spaces, and he employs an explanatory and detailed approach. With a systematic use of locally multiplicatively-convex algebras, the author presents theory regarding spectrum (local and global), projective and inductive limit algebras, the Gel'fand map, and tensor products; he also includes some current applications. LW

**Analysis, P.** *Lecture Notes in Mathematics-1334: Global Analysis—Studies and Applications III.* Ed: Yu. G. Borisovich, Yu. E. Gliklikh. Springer-Verlag, 1988, 331 pp, $28.60 (P). [ISBN: 0-387-50019-7] Translation of sixteen expository and selected research papers from the two 1987 issues of the Russian publication series "New Developments in Global Analysis." A sequence to *LNM-1108* (TR, December 1985) and *LNM-1214* (TR, June-July 1987). LAS

**Differential Geometry, P.** *Elliptic Operators, Topology and Asymptotic Methods.* John Roe. Pitman Res. Notes in Math. Ser., V. 179. Longman Scientific & Technical (US Distr: Wiley), 1988, 184 pp, $49.95 (P). [ISBN: 0-470-21095-8] A survey of material leading up to the heat equation proof of the Atiyah-Singer index theorem. Differential geometry and functional analysis, but not partial differential equations, form necessary background. Pricey paperback. GG

**Differential Geometry, T*(16-17).** *Manifolds, Tensor Analysis, and Applications.* R. Abraham, J.E. Marsden, T. Ratiu. Appl. Math. Sci., V. 75. Springer-Verlag, 1988, vii + 654 pp, $59.80. [ISBN: 0-387-96790-7] Interesting introduction to differential geometry and its applications. Written for a wide audience; requires only linear algebra and multivariable calculus. Topics include vector fields (including a dynamical system), both finite and infinite dimensional manifolds, differential forms, and integration. Applications come mainly from mechanics. Exercises. MR

**Geometry, S(16-17), P, L.** *Art Gallery Theorems and Algorithms.* Joseph O'Rourke. Intern. Ser. of Mono. on Comput. Sci., V. 3. Oxford U Pr, 1987, xiv + 282 pp, $45. [ISBN: 0-19-503965-3] The original Art Gallery Theorem states that $[n/3]$ guards are necessary and sufficient to guard a polygonal art gallery with $n$-walls. This extensive monograph discusses this result and other results of combinatorial and computational geometry which have a similar flavor. Includes references but not exercises. CEC

**Geometry, P.** *Geometry of Random Motion.* Ed: Rick Durrett, Mark A. Pinsky. Contemp. Math., V. 73. AMS, 1988, xii + 337 pp, $32 (P). [ISBN: 0-8218-5081-4] Twenty-six papers from the AMS conference at Cornell University in 1987. Study of diffusion processes on manifolds gives results in geometry (e.g., on harmonic functions) and in probability (e.g., stochastic calculus of variations). TH

**Topology, P.** *Fixed Point Theory and Its Applications.* Ed: R.F. Brown. Contemp. Math., V. 72. AMS, 1988, x + 268 pp, $28 (P). [ISBN: 0-8218-5080-6] Proceedings of a conference held at the International Congress of Mathematicians at Berkeley in August 1986. LC

**Optimization, T(16-17: 1, 2), L*.** *Discrete Optimization.* R. Gary Parker, Ronald L. Rardin. Comput. Sci. & Sci. Comput. Academic Pr, 1988, xi + 472 pp, $69.95. [ISBN: 0-12-545075-3] After an introductory chapter on computational complexity, presents both matroid and linear programming-based strategies for polynomial-time solvable problems. Also covers enumerative and polyhedral methods as well as some heuristic methods. Prerequisites for the text (including linear algebra, linear programming, and graph theory) are reviewed in appendices. AO

**Dynamical Systems, P.** *Pseudo-orbits of Contact Forms.* A. Bahri. Pitman Res. Notes in Math. Ser., V. 173. Longman Scientific & Technical (US Distr: Wiley), 1988, 296 pp, $66.95 (P). [ISBN: 0-582-01991-5] Motivated by the question of whether a contact vector field on an orientable manifold has a periodic orbit. This book investigates a related variational problem from the point of view of critical points at infinity, i.e., gradient lines which do not end up at critical points of the vector field. MR

**Control Theory, T(18), P, L.** *Lecture Notes in*

*Control and Information Sciences-108: Multivariable Control: A Graph-theoretic Approach.* K.J. Reinschke. Springer-Verlag, 1988, 274 pp, $36.60 (P). [ISBN: 0-387-18899-1] Written for engineers in control research and development, graduate students in control theory, and applied mathematicians. Seeks to overcome some of the difficulties of analyzing multivariable control systems via state-space analysis by using a graph theoretic approach. No exercises, but many examples. SM

**Control Theory, T(18: 1), P.** *Lecture Notes in Control and Information Sciences-109: Modern Aircraft Flight Control.* M. Vukobratović, R. Stojić. Springer-Verlag, 1988, viii + 288 pp, $38.80 (P). [ISBN: 0-387-19119-4] Considers a class of aircraft whose members can be modeled as rigid bodies with six degrees of freedom. Assumes prior knowledge of control theory and systems. No exercises. SM

**Probability, T(16: 1), S, L.** *Elementary Applications of Probability Theory.* Henry C. Tuckwell. Chapman & Hall, 1988, xiii + 225 pp, $59.95 (P). [ISBN: 0-412-30490-2] A nice book. The applications are mostly to the life sciences and engineering. Emphasizes reliability and simulation. Contains a particularly clear development of random walks and Markov chains. The writing is clear and inviting. The price is unfortunate. TAV

**Probability, P.** *Probabilities and Potential, C: Potential Theory for Discrete and Continuous Semigroups.* Claude Dellacherie, Paul-André Meyer. Transl: J. Norris. Math. Stud., V. 151. North-Holland (US Distr: Elsevier Science), 1988, xiv + 416 pp, $92 (P). [ISBN: 0-444-70386-1] This third of a planned four-volume work develops potential theory in a semigroup setting: excessive functions, reductions, sweeping, maximum principle. Includes discussions of capacitory methods and resolvant fields. TAV

**Probability, T(15-16: 1), L.** *An Introduction to Probabilistic Modeling.* Pierre Brémaud. Undergrad. Texts in Math. Springer-Verlag, 1988, xvi + 207 pp, $34. [ISBN: 0-387-96460-6] Topics covered include: basic concepts and models in probability, discrete random variables, random variables admitting a probability density, Gaussian vectors, the Poisson process, and convergence. Emphasis is on modeling and computation. Includes exercises with detailed solutions. Assumes a knowledge of the Riemann integral, sequences and series, and elementary matrix algebra. RH

**Probability, T(17-18: 2).** *Probability Theory: Independence, Interchangeability, Martingales, Second Editon.* Yuan Shih Chow, Henry Teicher. Texts in Stat. Springer-Verlag, 1988, xviii + 467 pp, $42. [ISBN: 0-387-96695-1] Covers the measure-theoretic foundations of probability theory. Includes independence, interchangeability, and martingales. Special emphasis is placed on stopping times. No knowledge of measure theory or probability is assumed. Includes numerous exercises. RH

**Probability, T(15-18: 1), L.** *Probability: Methods and Measurement.* Anthony O'Hagan. Chapman & Hall, 1988, xii + 291 pp, $30 (P); $65. [ISBN: 0-412-29540-7; 0-412-29530-X] An introductory text in probability which is presented from the personalist viewpoint. Emphasis is on how probabilities are derived in practice. Basic concepts and results are developed in an approach based on probability measurement. Includes exercises with solutions. Only prerequisites are basic algebra and calculus. RH

**Probability, T(2), P, L.** *Advanced Probability Theory.* Janos Galambos. Prob.: Pure & Appl., V. 3. Marcel Dekker, 1988, vii + 405 pp, $99.75. [ISBN: 0-8247-7873-1] Intended as a text for a graduate-level introductory course in probability theory or as a reference for modern probability theory. Assumes a knowledge of calculus and elementary complex analysis but not elementary probability theory. Covers the mathematical theory and applications, and includes almost all proofs. Topics include: expectation and integral, weak and strong convergence, transforms of distribution, infinite sequences of independent random variables, triangular arrays of independent random variables, independent and identically distributed random variables, conditional expectation, martingales, and stochastic processes. RH

**Stochastic Processes, S(18).** *Lecture Notes in Mathematics-1322: Stochastic Analysis.* Ed: M. Métivier, S. Watanabe. Springer-Verlag, 1988, vii + 197 pp, $20 (P). [ISBN: 0-387-19352-9] A collection of fourteen papers from a joint seminar on probability theory held June 16-19, 1987 at Ecole Normale Supérieure, Paris, sponsored by the France-Japan Cooperative Science Program. Ten papers in English; four in French. Main theme: stochastic analysis and applications to large scale systems. Individual topics: Malliavan calculus, infinite dimensional stochastic differential equations, stochastic partial differential equations, limit theorems for particle systems, diffusions in random environment, hydrodynamical models, etc. JJ

**Stochastic Processes, T*(18: 3), S(18), P**, L.** *An Introduction to the Theory of Point Processes.* D.J. Daley, D. Vere-Jones. Ser. in Stat. Springer-Verlag, 1988, xxi + 702 pp, $79. [ISBN: 0-387-96666-8] In this massive work (sixteen chapters, three appendices), the authors attempt to collect and organize the theory of point processes. The writing is fluid and the organization is compelling. Examples and exercises enough for a useful text; a huge bibliography. The appendices on measure, metric spaces, and martingales are very useful for the nonspecialist. TAV

**Stochastic Processes, S(18), P.** *Lecture Notes in Mathematics-1308: Two-Parameter Martingales and Their Quadratic Variation.* Peter Imkeller. Springer-Verlag, 1988, iv + 177 pp, $17.30 (P). [ISBN: 0-387-19233-6] Two-parameter martingales are examples of stochastic processes with two distinct "time" directions. While this book presents one- and two-parameter processes, extensions to more general $n$-parameter systems should be easily obtained. SM

**Stochastic Processes, P.** *Lecture Notes in Mathematics-1325: Stochastic Mechanics and Stochastic Processes.* Ed: A. Truman, I.M. Davies. Springer-Verlag, 1988, 220 pp, $20 (P). [ISBN: 0-387-50015-4] Eighteen papers, most from a 1986 conference at Swansea, Wales, United Kingdom. Central themes are large deviations and statistical mechanics, Nelson's stochastic mechanics and quantum diffusions, and simulations of Brownian motions and stochastic flows. TH

**Stochastic Processes, T?(17: 2), S(17-18), L*.** *Basic Stochastic Processes: The Mark Kac Lectures.* Reza Iranpour, Paul Chacon. Macmillan, 1988, xiv + 258 pp. [ISBN: 0-02-359820-4] A "short course" in stochastic processes based upon Kac's lectures. A non-measure approach that tries (and succeeds) to bring most of the basic ideas together. A nice reference book; some exercises, no bibliography. TAV

**Stochastic Processes, P*.** *Limit Theorems for Stochastic Processes.* Jean Jacod, Albert N. Shiryaev. Springer-Verlag, 1987, xvii + 601 pp, $98. [ISBN: 0-387-17882-1] The limit theorems in the title are those of weak convergence of measures in metric spaces. These are developed to provide a systematic approach to convergence in probability law for processes that are semi-martingales. Contains useful indices and a huge bibliography. TAV

**Stochastic Processes, P.** *Stochastic Equations for Complex Systems.* A.V. Skorohod. Math. & Its Applic. D Reidel (US Distr: Kluwer Academic), 1987, xvii + 175 pp, $69. [ISBN: 90-277-2408-3] The "complex" in the title implies large (or infinite) dimensional and topologically complex state spaces. The main tool developed is infinite systems of linear stochastic equations. Highly technical. TAV

**Stochastic Processes, T(18: 1), P.** *Probability, Random Processes, and Ergodic Properties.* Robert M. Gray. Springer-Verlag, 1987, xvi + 295 pp, $49.50. [ISBN: 0-387-96655-2] A terse, tightly written book to serve as a theoretical background for understanding the nature of (discrete) random processes. Special emphasis is on ergodicity and emergence of long-term time averages. Extensive bibliography. TAV

**Elementary Statistics, T, P, L*.** *Statistical Reasoning in Law and Public Policy, Volume 1: Statistical Concepts and Issues of Fairness.* Joseph L. Gastwirth. Stat. Modeling & Decision Sci. Academic Pr, 1988, xxii + 485 pp, $84.50. [ISBN: 0-12-277160-5] Introduction to basic statistical techniques (descriptive statistics, normal and binomial distributions, comparing proportions) with case studies of how these techniques have been used in legal cases (e.g., equal employment, health risks, jury selection). Aims more at developing informed common sense than expertise in carrying out statistical analyses. LAS

**Elementary Statistics, T(13-15).** *Statistical Methods for the Social and Behavioral Sciences.* Leonard A. Marascuilo, Ronald C. Serlin. WH Freeman, 1988, xxvi + 804 pp, $44.95. [ISBN: 0-7167-1824-3]

**Elementary Statistics, T(12-14), S, L, C.** *Data Analysis.* IBM PC. North Carolina School of Science. NCTM, 1988, vi + 132 pp, $10 (P). [ISBN: 0-87353-258-9] Part of a senior course *Introduction to College Mathematics* developed at the North Carolina School of Science and Mathematics. Three chapters (single variable, bivariate curve fitting, linear regression) introduce basic topics with extensive real data-rich examples. Software disk provides tools to create, save, and analyze data sets. Topics and approach match well the new NCTM recommendation for data analysis in the core of school mathematics. LAS

**Computational Statistics, P.** *COMPSTAT: Proceedings in Computational Statistics.* Ed: D. Edwards, N.E. Raun. Physica-Verlag (US Distr: Springer-Verlag), 1988, xiv + 451 pp, $68.40 (P). [ISBN: 0-387-91336-X] Sixty select papers from the 1988 symposium in computational statistics held in Copenhagen. Topics include the application of artificial intelligence to statistics, nonparametric estimation, graphical techniques, algorithms, projection pursuit, and other computationally-intensive statistical methods. TH

**Statistics, P*.** *Recent Developments in Clustering and Data Analysis.* Ed: Edwin Diday, *et al.* Academic Pr, 1988, xv + 452 pp, $49.50. [ISBN: 0-12-215485-1] Proceedings of the Japanese-French Scientific Seminar held in March 1987 in Tokyo. Thirty-five papers in the areas of clustering and multidimensional data analysis, discussing methods developed in France and Japan which are relatively unknown in English-speaking countries. RSK

**Statistics, S(15-17) L.** *Probability Theory and Mathematical Statistics.* G. Klimov. Transl: Oleg Efimov. MIR (US Distr: Imported Pub), 1986, 334 pp, $9.95. [ISBN: 0-8285-3214-1]

**Statistics, T(17-18: 1).** *Extreme Value Theory in Engineering.* Enrique Castillo. Stat. Modeling & Decis. Sci. Academic Pr, 1988, xv + 389 pp, $59.95. [ISBN: 0-12-163475-2] Makes available to practicing engineers and scientists many recent advances in extreme value theory. Gives methods for applying newest theory through many examples with real engineering data sets. Good motivation of need for extreme value theory in engineering by examples (extreme wind speed analysis in structural engineering). Topics addressed: order statistics, asymptotic distributions of maxima and minima, Weibull, Gumbel and Frechet distributions, limit distributions, and multivariate and regression problems related to extremes. JJ

**Statistics, T, P, L.** *Foundations of Statistics.* D.G. Rees. Chapman & Hall, 1987, xv + 543 pp, $27.50 (P). [ISBN: 0-412-28560-6]

**Statistics, P*.** *Order Restricted Statistical Inference.* Tim Robertson, F.T. Wright, R.L. Dykstra. Prob. & Math. Stat. Wiley, 1988, xix + 521 pp, $59.95. [ISBN: 0-471-91787-7] Up-to-date presentation inspired by the 1972 book *Statistical Inference under Order Restrictions* by Barlow, *et al.* (TR, Oc-

tober 1973), emphasizing procedures based on the likelihood principle. Comprehensive treatment of statistical inference under inequality constraints. Extensive set of references. RSK

**Statistics, T(15-16) P, L.** *Linear Estimation and Design of Experiments.* D.D. Joshi. Wiley, 1987, xv + 288 pp, $24.95. [ISBN: 0-470-20740-X]

**Statistics, T(15-16), S, C.** *TIMESLAB: A Time Series Analysis Laboratory.* IBM PC. H. Joseph Newton. Wadsworth, 1988, xv + 623 pp, (P). [ISBN: 0-534-09198-9] A text with accompanying software (for IBM compatible color PC systems) on fundamental theory, methods, and algorithms for time series, combining time domain and frequency domain approaches. Three chapters treat univariate time series; one analyzes bivariate series. Appendices provide necessary mathematical and statistical topics (matrix theory, fast Fourier transform, regression); a user's guide to the software package TIMESLAB; detailed descriptions of commands; text of macros, data sets, and examples. Software comes on two disks, not copy protected, with careful instructions for installation on two drive or hard disk systems (shared or private). Software can be used either in command mode or in program mode with macros, many of which are provided on disk. LAS

**Statistics, P, L.** *Robust Regression and Outlier Detection.* Peter J. Rousseeuw, Annick M. Leroy. Appl. Prob. & Stat. Wiley, 1987, xiv + 329 pp, $36.95. [ISBN: 0-471-85233-3]

**Statistics, S*(16-18), P*.** *Lecture Notes in Statistics-45: Statistical Information and Likelihood: A Collection of Critical Essays by Dr. D. Basu.* Ed: J.K. Ghosh. Springer-Verlag, 1988, xviii + 365 pp, $39.70 (P). [ISBN: 0-387-96751-6] Intriguing collection of nineteen articles by Basu on the foundations of statistical inference. Three include discussions by eminent statisticians with replies by Basu, and many contain author's notes which add an interesting personal touch. A master of the counterexample, Basu concludes "that most of the statistical methods that I learned from pioneers like Karl Pearson, Ronald Fisher and Jerzy Neyman and survey practitioners like Morris Hanson, P.C. Mahalanobis and Frank Yates are logically untenable." RSK

**Elementary Computer Science, S(13).** *Number Problem Visual Masters with BASIC Program Solutions.* Donald D. Spencer. Camelot, 1988, 96 pp, $13.95 (P) [ISBN: 0-89218-156-7]; *Number Problem Visual Masters with PASCAL Program Solutions,* Camelot, 1988, 96 pp, $13.95 (P). [ISBN: 0-89218-157-5] Collection of visual masters for transparencies or handouts involving problem solving with BASIC and PASCAL. Includes program solution to problems. Most problems involve elementary number theoretic properties of integers. SB

**Elementary Computer Science, S(13).** *Math Problems Visual Masters with BASIC Program Solutions.* Donald D. Spencer. Camelot, 1988, 96 pp, $13.95 (P). [ISBN: 0-89218-154-0] A collection of forty-five problem masters and forty-five associated BASIC program solution masters for overhead transparencies. Appropriate for use as examples in a very elementary BASIC programming course. RH

**Languages, S(15-18), L.** *The Turing Programming Language: Design and Definition.* Richard C. Holt, *et al.* Prentice-Hall, 1988, viii + 325 pp, (P). [ISBN: 0-13-933136-0] Presents a detailed case study of the design of the Turing programming language. Contains the language's formal definition, the Turing report, the language's formal semantics, six appendices, bibliography, and index. RJA

**Theory of Computation, T(16-17: 1), S, L.** *Reflexive Structures: An Introduction to Computability Theory.* Luis E. Sanchis. Springer-Verlag, 1988, xii + 233 pp, $39.95. [ISBN: 0-387-96728-1] An introductory text in the general theory of computability which develops closure, reflexivity, enumeration, and hyperenumeration properties concretely in terms of natural numbers and function on them, avoiding abstract axiomatic domains. Reflexivity is capitalized upon to support an economical presentation. Nearly 300 exercises at ends of sections. RB

**Artificial Intelligence, T(16-18: 1), S.** *Introduction to Expert Systems.* Peter Jackson. Intern. Comput. Sci. Ser. Addison-Wesley, 1986, ix + 246 pp, (P). [ISBN: 0-201-14223-6] A book on artificial intelligence built around expert systems as the most important application of artificial intelligence. Three formalisms (production rules, frames, predicate logic) for organizing expert systems are presented. A number of well-known expert systems are discussed and related to one or more of the formalisms. Subsequent topics are knowledge acquisition, learning, explanation, and expert system tools. Some exercises; bibliography; index. RJA

**Artificial Intelligence, P.** *Lecture Notes in Computer Science-286: Uncertainty in Knowledge-Based Systems.* Ed: B. Bouchon, R.R. Yager. Springer-Verlag, 1987, viii + 405 pp, $33.30 (P). [ISBN: 0-387-18579-8]

**Artificial Intelligence, T(16-18: 1), S, P.** *Structured Induction in Expert Systems.* Alen D. Shapiro. Addison-Wesley, 1987, xiii + 134 pp. [ISBN: 0-201-17813-3] Report on an experiment to see if machine learning can alleviate the domain knowledge acquisition bottleneck in the development of expert systems. Chess endgames were used in the experiment. The central issue is the use of mechanized inductive learning to build expert systems from a collection of examples. References; appendices; index. RJA

**Artificial Intelligence, T(15-18: 1) S*.** *Algorithmic Methods for Artificial Intelligence.* Michael Griffiths, Carol Palissier. Chapman & Hall, 1987, 144 pp, $19.95 (P); $34.50. [ISBN: 0-412-01541-2; 0-412-01541-5] Presents structures and algorithms commonly used in artificial intelligence programming. Organization is succinct and often built around examples. Includes an historical introduction, application to expert systems, and a chapter on LISP and PROLOG. Some exercises, references, index. RJA

**Computer Science, P.** *Lecture Notes in Computer*

*Science-298: Mathematical Foundations of Programming Language Semantics.* Ed: M. Main, *et al.* Springer-Verlag, 1988, viii + 637 pp, $50.90 (P). [ISBN: 0-387-19020-1] Selection of papers from the third workshop on the mathematical foundations of programming language semantics, held at Tulane University in April, 1987. RJA

**Computer Science, P.** *Lecture Notes in Computer Science-301: Pattern Recognition.* Springer-Verlag, 1988, vii + 668 pp, $67.90 (P). [ISBN: 0-387-19036-8]

**Computer Science, P, L.** *Lecture Notes in Computer Science-310: 9th International Conference on Automated Deduction.* Ed: E. Lusk, R. Overbeek. Springer-Verlag, 1988, x + 775 pp, $63.60 (P). [ISBN: 0-387-19343-X] Commemorates the twenty-fifth anniversary of the discovery of the resolution principle. Includes papers on theorem proving, logic programming, unification, deductive databases, term rewriting, ATP for non-standard logics, and program verification. RJA

**Computer Science, P.** *Lecture Notes in Computer Science-315: Logic Programming '87.* Ed: K. Furukawa, H. Tanaka, T. Fujisaki. Springer-Verlag, 1988, vi + 327 pp, $25.70 (P). [ISBN: 0-387-19426-6] Contains papers presented at the Sixth Logic Programming Conference. RJA

**Computer Science, T(15-17: 1), S.** *Computer Simulation and Computer Algebra: Lectures for Beginners.* D. Stauffer, *et al.* Springer-Verlag, 1988, ix + 155 pp, $19.80 (P). [ISBN: 0-387-18909-2] A friendly, enthusiastic introduction to numerical and algebraic computational methods for mathematical physics. The first half of the book illustrates numerical techniques (Runge-Kutta methods, Monte Carlo simulations, etc.) implemented in FORTRAN, applied to problems in classical physics and statistical mechanics. The last half is a fairly detailed introduction to the computer algebra system REDUCE. An appendix covers elementary techniques and philosophy of FORTRAN programming. PZ

**Computer Science, P.** *Logic Programming: Proceedings of the Fourth International Conference.* Ed: Jean-Louis Lassez. Logic Programming. MIT Pr, 1987, $47.50 set (P). [ISBN: 0-262-12125-5] *Volume 1,* xiv + 555 pp; *Volume 2,* xiv + 500 pp. Papers grouped under the headings: Warren Abstract Machine; Databases; Constraints; Parallelism; Implementation Issues; Language Issues; Applications; Program Analysis; Concurrent Languages; Invited Talks. RJA

**Computer Science, P.** *Lecture Notes in Computer Science-326: ICDT '88.* Ed: M. Gyssens, J. Paredaens, D. Van Gucht. Springer-Verlag, 1988, vi + 409 pp, $31.40 (P). [ISBN: 0-387-50171-1] Proceedings of the second biennial European-sponsored International Conference on Database Theory, Bruges, Belgium, August 1988. Three invited talks, 23 selected papers on dynamic aspects of databases, logic and deductive databases, complexity and optimization, object-oriented databases, data models and query languages, other topics. RB

**Computer Science, P.** *Lecture Notes in Computer Science-322: ECOOP '88: European Conference on Object-Oriented Programming.* Ed: S. Gjessing, K. Nygaard. Springer-Verlag, 1988, vi + 410 pp, $31.40 (P). [ISBN: 0-387-50053-7]

**Applications, S*(16-17), P, L.** *Lecture Notes in Computer Science-325: Modern Cryptology: A Tutorial.* Gilles Brassard. Springer-Verlag, 1988, vi + 107 pp, $15.40 (P). [ISBN: 0-387-96842-3] Notes from a 1987 IEEE tutorial giving a self-contained survey of recent advances in cryptology which is both elementary yet up-to-date. Covers secret key systems, public key systems, applications (e.g., authentication, coin flipping by telephone, protection of privacy), and quantum cryptography in which information is encoded in the quantum states of photons. LAS

**Applications, T(13-16), S**.** *Mathematical Modelling.* J.N. Kapur. Wiley, 1988, xi + 259 pp, $24.95. [ISBN: 0-470-20088-X] A potpourri of modeling applications suitable for presentation in freshman through senior level courses. Contents grouped according to techniques: ordinary differential equations, systems, difference equations, partial differential equations, graphs, integral-differential equations, calculus of variations, and maximum-entropy principle. Could serve as a text for a modeling course, but because of wide range of topics, probably best suited for a reference. Exercises. MR

**Applications (Biological Science), P.** *Lecture Notes in Biomathematics-74: Computer Simulation in Cell Radiobiology.* Andrej Yu. Yakovlev, Aleksandr V. Zorin. Springer-Verlag, 1988, vi + 133 pp, $16.30 (P). [ISBN: 0-387-19457-6] Demonstrates possible ways of using simulation for modeling cell kinetics, with emphasis on the effects of cell radiobiology. Includes concrete applications of the model. Should be of interest to biologists, biomathematicians, and computer scientists interested in simulation modeling as a tool for exploring biological processes. RH

**Applications (Communication Theory), T (18), P.** *Multiplicative Complexity, Convolution, and the DFT.* Michael T. Heideman. Springer-Verlag, 1988, viii + 155 pp, $32. [ISBN: 0-387-96810-5] Theory of multiplicative complexity, mainly as applied to digital signal processing computations. Studies systems of polynomial multiplications, polynomial products with constraints, and the discrete Fourier transform. Appendices, complexities of multidimensional cyclic convolutions, and one-dimensional DFT. Problems. LC

**Applications (Economics), P.** *Lecture Notes in Mathematics-1330: Mathematical Economics.* Ed: A. Ambrosetti, F. Gori, R. Lucchetti. Springer-Verlag, 1988, vii + 137 pp, $13.10 (P). [ISBN: 0-387-50003-0] Texts of four survey courses presented at the CIME Session on "Mathematical Economics" held at Villa La Querceta in Montecatini Terme, Italy, from June 25 to July 3, 1986. The four texts are "Some Variational Methods Arising from Mathematical Economics" by I. Ekeland, "Four Lectures

on the Differentiable Approach to General Equilibrium Theory" by A. Mas-Colell, "Dynamic General Equilibrium Models" by J. Scheinkman, and "Topics in Non Cooperative Game Theory" by S. Zamir. RH

**Applications (Economics), T(17: 1), S, L.** *Foundations of the Theory of General Equilibrium.* Yves Balasko. Econ. Theory, Econometrics, & Math. Econ. Academic Pr, 1987, xiii + 285 pp, $39.95. [ISBN: 0-12-076975-1] Differential topology of equilibria for pure exchange economics. Requires an undergraduate or early graduate level mathematics background and an introductory course in microeconomics. Focus on geometry of equilibria; little on stability. GG

**Applications (Engineering), P.** *Studies in Nonlinear Aeroelasticity.* Earl H. Dowell, Marat Ilgamov. Springer-Verlag, 1988, xvi + 455 pp, $68. [ISBN: 0-387-96791-5] The traditional linear models don't explain everything about stability and motion of flexible structures in a flowing fluid. A good book to read on a long airplane flight. BC

**Applications (Engineering), S(18), P.** *Analysis and Estimation of Stochastic Mechanical Systems.* Ed: W. Schiehlen, W. Wedig. CISM Courses & Lect., No. 303. Springer-Verlag, 1988, 350 pp, $41.10 (P). [ISBN: 0-387-82058-2] A collection of eight lectures and tutorials presented during a course at the International Centre for Mechanical Sciences in Udine in summer 1987. Summarizes the latest developments in stochastic analysis and estimation with theoretical applications to practical problems in mechanical systems. Topics include random vibrations of systems, stochastic modeling of fatigue damage and crack growth, modeling of vehicle-road systems, and analysis of nonlinear systems. Of interest to engineers with advanced degrees in engineering mathematics. JJ

**Applications (Fluid Dynamics), P.** *Numerical Solution of the Shallow-Water Equations.* F.W. Wubs. CWI Tract, V.49. Mathematisch Centrum, 1988, iv + 115 pp, Dfl. 17.80 (P). [ISBN: 90-6196-349-4] Appropriately published in The Netherlands, most of the book is a mixture of discrete numerical methods and, to lesser extent, of computer code used to model shallow-water flow. Includes short treatment of stabilization and smoothing. GG

**Applications (Fluid Dynamics), P, L.** *Stably Stratified Flow and Dense Gas Dispersion.* Ed: J.S. Puttock. IMA Conf. Ser., V. 15. Clarendon Pr, 1988, xiv + 430 pp, $89. [ISBN: 0-19-853615-1] Proceedings of an April 1986 conference held in Chester, England, emphasizing the atmospheric dispersion of gases which are denser than air—a topic of vital importance in predicting the consequences of accidental release of many toxic or flammable gases. The third in a series of conferences on environmental mathematics sponsored by the Institute of Mathematics and its Applications. LAS

**Applications (Information Theory), T(18), P.** *Information Mechanics: Transformation of Information in Management, Command, Control and Communication.* Brian Conolly, John G. Pierce. Math. & Its Applic. Halsted Pr, 1988, 174 pp, $44.95. [ISBN: 0-470-21136-9] Uses information theory to address concerns in military command, control, and communications. Other mathematical formalisms proposed for the modeling of $C^3$ include control theory, fuzzy set theory, and catastrophe theory. This work uses information theory to address questions like the identification of true and/or false targets, management of information channels, and the modification of the classical Lancaster equations to incorporate the effects of information. SM

**Applications (Physical Science), S(15-17), P, L.** *Lecture Notes in Earth Sciences-18: Numerical Geology.* N.M.S. Rock. Springer-Verlag, 1988, xi + 427 pp, $39.50 (P). [ISBN: 0-387-50070-7] A vast compendium of techniques, sources, and examples of the use of computers and statistics in geology. Not so much an introduction as a handbook, these notes provide key facts with references to diverse statistical, numerical, and graphing techniques. Includes a 2000-item "selective" bibliography and a 45 page glossary-index. LAS

**Applications (Physics), P.** *Conformal Field Theory and Solvable Lattice Models.* Ed: M. Jimbo, T. Miwa, A. Tsuchiya. Adv. Stud. in Pure Math., V. 16. Academic Pr, 1988, 426 pp, $59.50. [ISBN: 0-12-385340-0] Nine papers in mathematical physics. Topics include conformal field theory, solvable lattice models, affine and Virasoro algebras, and KP equations. MR

**Applications (Physics), S*(17-18), P, L.** *Quantum Probability.* Stanley P. Gudder. Prob. & Math. Stat. Academic Pr, 1988, xii + 316 pp, $49.50. [ISBN: 0-12-305340-4] Well-written generalization of classical probability theory needed for several approaches to quantum mechanics, in particular those of operational statistics and of the path-integral formalism. Largely self-contained on the whole, even in introductory chapters on classical probability and classical quantum mechanics. GG

### Reviewers

RJA: Richard J. Allen, St. Olaf; DFA: David F. Appleyard, Carleton; SB: Steve Benson, St. Olaf; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; CE: Christopher Ennis, Carleton; SG: Steven Galovich, Carleton; GG: George Gilbert, St. Olaf; BH: Bruce Hanson, St. Olaf; RH: Rhonda Hatcher, St. Olaf; TH: Timothy Hesterberg, St. Olaf; JJ: Jason Jones, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; GN: Gail Nelson, Carleton; AO: Arnold Ostebee, St. Olaf; SP: Samuel Patterson, Carleton; MR: Matthew Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MS: Myriam Steinback, Macalester; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; LW: Lynne Walling, St. Olaf; PZ: Paul Zorn, St. Olaf.

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 96, Number 4 April 1989

# Contents

(ISSN 0002-9890)

ARTICLES

On Groups in Which Every Element Has Finite Order . . . . . . . . . NARAIN GUPTA 297

The Shortest Planar Arc of Width 1 . . . . . . . . . . . ANI ADHIKARI AND JIM PITMAN 309

The Use of Full Covers in Real Analysis . . . . . . . . . . . . . . . MICHAEL W. BOTSKO 328

UNSOLVED PROBLEMS

What Are the Laws of Greed? . . . . . . . . . . . . . . . . . . . . . . . . . . . . . JIM PROPP 334

NOTES

Fair Dice . . . . . . . . . . . . . . . . . . . . . . . PERSI DIACONIS AND JOSEPH B. KELLER 337

A Remark on Euclid's Proof of the Infinitude of Primes . . . . JOHN B. COSGRAVE 339

An Alternate Proof of the Continuity of the Roots of a Polynomial . . . . . . . . . . . . . . FELIPE CUCKER AND ANTONIO G. CORBALAN 342

THE TEACHING OF MATHEMATICS

Taylor's Theorem Using the Generalized Riemann Integral . . . . H. B. THOMPSON 346

The Effect of Prior Calculus Experience on "Introductory" College Calculus . . . . . . . . . . . . . . . . . . . . MARTHA B. BURTON 350

Simple Inequalities and Old Limits . . . . . . . . . . . . . . . . . . . . CHUNG-LIE WANG 354

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 356

Advanced Problems and Solutions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 366

REVIEWS

Geometries and Groups by V. V. Nikulin and I. R. Shafarevich . . . . . . . . . HEINRICH W. GUGGENHEIMER 370

Mathematical Cryptology for Computer Scientists and Mathematicians by Wayne Patterson; A Course in Number Theory and Cryptography by Neal Koblitz . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . ALAN G. KONHEIM 374

Invitation to Complex Analysis by Ralph P. Boas . . . . . . . . . . GEORGE PIRANIAN 376

Principles of Computer Science by M. Sandra Carberry, A. Toni Cohen, and Hatem M. Khalil . . . . . . . . . . . . . . . . . . . . . . . . . . STANLEY E. SELTZER 378

Writing Mathematics Well by Leonard Gillman . . . . . . . . . . . . . . . PETER D. LAX 380

MISCELLANEOUS

P.G.O.M. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . ALLEN J. SCHWENK 382

TELEGRAPHIC REVIEWS . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 383

# NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to ANITA E. SOLOW, Department of Mathematics, Grinnell College, Grinnell, IA 50112; in geometry or topology to DENNIS DETURCK, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems (three copies) and solutions (two copies), both elementary and advanced to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, University of Calgary, Alberta, Canada T2N 1N4




Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: MS. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $26 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $29. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $90 per year.





Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Assocation of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

# On Groups in Which Every Element Has Finite Order

NARAIN GUPTA, *University of Manitoba*

NARAIN GUPTA is Professor of Mathematics at the University of Manitoba, Canada, where he has taught since 1967. He received his M.A. and LL.B. (1959) degrees from Aligarh University, India, and his Ph.D. (1965) from the Australian National University, Canberra, where he served as a lecturer between 1965–1967. His current research interests are in the areas of combinatorial group theory, and he has published several papers on Burnside groups, Fox subgroups, and dimension subgroups. Narain Gupta is a Fellow of the Royal Society of Canada and a recipient of a Killam Research Fellowship (1965–1967) from the Canada Council.

*A still undecided point in the theory of discontinuous groups is whether the order of a group may be not finite, while the order of every operation it contains is finite. A special form of this question may be stated as follows: Let $A_1, A_2, \ldots, A_m$ be a set of independent operations, finite in number, and suppose that they satisfy the system of relations given by $S^n = 1$, where n is a given finite integer, while S represents in turn any and every operation which can be generated from the m given operations A. Is the group thus defined one of finite order, and if so what is its order?*

Burnside (1902).

**1.** In the modern terminology a group in which every element has finite order is said to be *periodic* (or *torsion*). There are two parts to the question raised by Burnside, and the first part has been interpreted as: Are finitely generated periodic groups finite? [The additive group of the infinite-dimensional vector space over the field of two elements is an infinite periodic group, so the assumption that the group be finitely generated is essential.] I shall refer to this question as the *General Burnside Problem*. To state the second part of the question, let $m, n$ be natural numbers and $G$ be a group which can be generated by $m$ elements and in which $g^n = 1$ for all $g \in G$. Then $G$ is a homomorphic image of a universal group $B(m, n)$ known as the *free m-generator Burnside group of exponent n* which can be generated by $m$ elements $x_1, \ldots, x_m$, and in which every element has order dividing $n$, and every other relation in $B(m, n)$ is a consequence of these power relations. In other words, $B(m, n)$ admits a *presentation* of the form:

$$B(m, n) = \langle x_1, \ldots, x_m;\ w^n = 1 \text{ for all words } w \text{ in the alphabet } x_1, \ldots, x_m \rangle.$$

The second part of Burnside's question reads: *Is $B(m, n)$ finite and if so, what is its order*? I shall refer to the first half of this question as the *Burnside Problem* and the second half as the *Order Problem*. Clearly $B(m, 1)$ is the trivial group and $B(1, n)$ is cyclic of order $n$. For a better understanding of the complexity of the Burnside Problem for $m, n > 1$, consider the Burnside group $B(4, n)$ for some $n > 1$. I begin by listing an infinite sequence of words in $B(4, n)$ as follows:

$$w_1 = x_1x_2x_3x_4;$$
$$w_2 = x_1x_2x_3x_4x_1x_4x_3x_2;$$
$$w_3 = x_1x_2x_3x_4x_1x_4x_3x_2x_1x_2x_3x_2x_1x_4x_3x_4.$$

Having defined $w_k$, a word of length $2^{k+1}$, one can regard

$$w_k = w_{k1}w_{k2}w_{k3}w_{k4}$$

as a product of its four quarters of length $2^{k-1}$ each, and define

$$w_{k+1} = w_k w_{k1} w_{k4} w_{k3} w_{k2}.$$

It can be shown that $w_k$ has no subword of the form $uu$ [Dean (1965)] [two-letter infinite sequences without subwords of the form $uuu$ were constructed by Arshon (1937) and Morse and Hedlund (1944)]. Thus, even $B(4,2)$ appears to be infinite. However, if $B(4, n)$ were finite then all of the words $w_k$ would have reduced to a finite subset of $B(4, n)$ as a consequence of the *law* $w^n = 1$. It is this consequential aspect of the law $w^n = 1$ which makes Burnside's question a challenging problem.

The finiteness of $B(m,2)$ is easy:

$$\begin{aligned} x_i x_j &= x_i x_j (x_j x_i x_j x_i) \\ &= x_i (x_j x_j) x_i x_j x_i = x_i x_i x_j x_i = x_j x_i, \end{aligned}$$

and, therefore, $B(m,2)$ reduces to a finite set $\{x_1^{\alpha(1)} x_2^{\alpha(2)} \cdots x_m^{\alpha(m)}, \alpha(i) \in \{0,1\}\}$ consisting of at most $2^m$ elements. This solves the Burnside Problem for $B(m,2)$. To solve its Order Problem observe that the additive group of the $m$-dimensional space over the two element field $Z_2$ is a homomorphic image of $B(m,2)$ and has order precisely $2^m$. Hence $B(m,2)$ is an elementary abelian 2-group and $|B(m,2)| = 2^m$.

Burnside offered solutions in some special cases of the problems and proved:

$$\text{(i) } |B(2,3)| = 27; \qquad \text{(ii) } |B(m,3)| < \infty; \qquad \text{(iii) } |B(2,4)| \leqslant 2^{12}.$$

[Towards the general problem, Burnside (1905) also proved that finitely generated periodic groups of matrices (over some field) are finite.] The Order Problem for $B(m,3)$ was resolved by Levi and van der Waerden (1933). They proved that every element of $B(m,3)$ can be uniquely reduced to the *normal form*

$$\prod_i x_i^{\alpha(i)} \prod_{i<j} [x_i, x_j]^{\alpha(i,j)} \prod_{i<j<k} [x_i, x_j, x_k]^{\alpha(i,j,k)},$$

where $[x_i, x_j]$ $(= x_i^{-1} x_j^{-1} x_i x_j)$ and $[x_i, x_j, x_k]$ $(= [[x_i, x_j], x_k])$ are standard *left-normed commutators* and $\alpha(i)$, $\alpha(i, j)$, $\alpha(i, j, k) \in \{0,1,2\}$. This yields

$$|B(m,3)| = 3^{\binom{m}{1}+\binom{m}{2}+\binom{m}{3}}$$

[earlier Burnside had shown that $|B(m,3)| \leqslant 3^{f(m)}$, $f(m) = 2^m - 1$]. I give a brief outline of how the above normal form can be achieved. It will be convenient to set $G = B(m,3)$ and use $a, b, c, \ldots$ for its elements. First, prove that $G$ satisfies the second Engel condition, namely, $[a, b, b] = 1$ for all $a, b$. Second, prove that a second Engel group satisfies the laws $[a, b, c, d] = 1$ and $[a, b, c] = [b, c, a] = [c, a, b]$ for all $a, b, c, d$. Finally, construct a 3-generator group of exponent 3 in which $[a, b, c] \neq 1$ for some $a, b, c$. The above normal form is an easy consequence of these results and the *collecting process*: $ab = ba[a, b]$. See Magnus, Karrass and Solitar (1966) for details.

The next important contribution towards the Burnside Problem came in 1940 with the appearance of a paper by Sanov, who proved that $B(m,4)$ is finite. The proof consists in showing that if $H$ is a finite group of exponent 4 and if $G$ is generated by $H$ together with an element $x$, and if, moreover, $x^2 \in H$ and $g^4 = 1$

for all $g \in G$, then $G$ is also finite. From this the finiteness of $B(m,4)$ follows by an easy induction on $m \geqslant 1$. For details see Magnus et al. The Order Problem for $B(m,4)$ has not yet been resolved [I will return to the current status of this problem in Section 3.]

While the finiteness of $B(2,5)$ remains unresolved, in 1958 Marshall Hall, Jr. proved that the Burnside groups $B(m,6)$ are finite of order given by the formula

$$|B(m,6)| = 2^{\alpha}3^{\binom{\beta}{1}+\binom{\beta}{2}+\binom{\beta}{3}},$$

where $\alpha = 1 + (m-1)3^{\binom{m}{1}+\binom{m}{2}+\binom{m}{3}}$ and $\beta = 1 + (m-1)2^m$. [Hall's work was encouraged by a fundamental paper of Philip Hall and Graham Higman (1956) in which, among other things, the above order was predicted in case $B(m,6)$ turned out to be finite.]

In the meantime, some progress was being reported towards the solution of the so-called *Restricted Burnside Problem* for $B(m,n)$ [which seems to have been initiated by Grün in (1940)]: *Is there a bound on the orders of finite quotients of* $B(m,n)$? [The term "restricted Burnside problem" was introduced by Magnus (1950).] The Hall and Higman paper [together with the Classification Theorem for finite simple groups] reduces the restricted problem for $B(m,n)$ *for all* $m$ to the restricted problem for $B(m,q)$ *for all* $m$, where $q$ ranges over all prime-power divisors of $n$. For $n = p^k$, $p$ prime, finite quotients of $B(m,p^k)$ are nilpotent and the Restricted Burnside Problem takes the form: *Does the lower central series of* $B(m,p^k)$ *become stationary after finitely many steps*?

[For any group $G$, let $\gamma_n(G)$, $n \geqslant 1$, denote the $n$th term of its lower central series defined as follows: $\gamma_1(G) = G$ and for $i \geqslant 1$, $\gamma_{i+1}(G) = [\gamma_i(G), G] = \langle [a,b] | a \in \gamma_i(G), b \in G\rangle$. The group $G$ is said to be *nilpotent of class* $c$ if $\gamma_{c+1}(G) = 1$ for some $c \geqslant 0$. If, in addition, $\gamma_c(G) \neq 1$ then we say that $G$ is *nilpotent of class precisely* $c$. The property $[\gamma_i(G), \gamma_j(G)] \leqslant \gamma_{i+j}(G)$, $i, j \geqslant 1$, allows us an access to the so-called *Associated Lie Ring* $L(G) = \sum_{i\geqslant 1} \oplus\, \gamma_i(G)/\gamma_{i+1}(G)$, with the Lie operation [,] defined for homogeneous elements by $[a\gamma_{i+1}(G), b\gamma_{j+1}(G)] = [a,b]\gamma_{i+j+1}(G)$ and extended to arbitrary elements of $L(G)$ by additivity. The anticommutativity of $L(G)$, i.e., $[x,y] = -[\gamma, y]$, follows from the commutator identity $[a,b] = [b,a]^{-1}$. To verify the Jacobi identity: $[x,y,z] + [y,z,x] + [z,x,y] = 0$, one simply uses the group identity: $[a,b,c^a][c,a,b^c][b,c,a^b] = 1$, where $g^h$ means $h^{-1}gh$. We define the lower central series of $L(G)$ analogously: $L^1(G) = L(G)$, $L^{i+1}(G) = [L^i(G), L(G)]$, $i \geqslant 1$; alternatively, $L^{i+1}(G)$ is the additive subgroup generated by all left-normed Lie elements $[a_1\gamma_2(G), \ldots, a_{i+1}\gamma_2(G)]$. It is a simple matter to verify that $\gamma_n(G) = \gamma_{n+1}(G)$ if and only if $L^n(G) = 0$, Lie rings with this condition being called *nilpotent*. For details we refer to the book by Marshall Hall, Jr. (1959).]

The Lie ring techniques developed by Magnus in the 1930's turn out to be particularly suitable for the restricted problem for prime exponent. Preliminary work concerning the connection between the restricted problem for prime exponent and local nilpotency of certain Lie rings of prime characteristic was done by Magnus, Sanov and others in the early 1950's [see also Lyndon (1954), (1955) for a power-series approach]. It turns out that the associated Lie ring $L(B(m,p))$ has

characteristic $p$ and satisfies the $(p-1)$th Engel condition $[x, {}_{p-1}y] = 0$ for all $x, y$. [See, for instance, Hall (1959).] [Here $[x, {}_0y] = x$ and $[x, {}_{i+1}y] = [[x, {}_iy], y]$ for $i \geqslant 0$.] It follows that $L(B(m, p))$ is a homomorphic image of an $m$-generator Lie algebra $\mathscr{L}_p(m, p-1)$, over the field of $p$ elements, which is free with respect to the $(p-1)$th Engel condition, so that the nilpotence of $L(B(m, p)$ would follow if $\mathscr{L}_p(m, p-1)$ were to be nilpotent, and which, in turn, would yield an affirmative answer to the restricted problem for $B(m, p)$. The nilpotency of $\mathscr{L}_5(m, 4)$ was proved by Kostrikin (1955) for $m = 2$ and by Higman (1956) for all $m$. In 1959, Kostrikin proved the nilpotency of $\mathscr{L}_p(m, \mathrm{p} - 1)$ for all primes $p$ and all $m$, and thereby the restricted Burnside problem was resolved in the affirmative for groups of prime exponent.

Kostrikin's initial proof was somewhat opaque and he published a supplementary paper in 1979 in which he provided further details of his important techniques. However, Adian (1984) pointed out that some more details were necessary. A fully detailed proof is now available in Kostrikin's book (1986). An attempt to prove the finiteness of $B(2,5)$ has been made by Hall and Sims (1982). They prove that $B(2,5)$ has a normal subgroup of index $5^{10}$ which, they feel, should be an elementary abelian group of order $5^{24}$ [I add that the correct order of the largest finite quotient of $B(2,5)$ is known to be $5^{34}$; see Havas, Wall, and Wamsley (1974)]. It may be of interest to mention a most recent result of Adian and Repin (1986) that the nilpotency class of $\mathscr{L}_p(2, p-1)$ is at least $2^{(p-1)/15}$.

In 1959 some eyebrows were raised when P. S. Novikov in Moscow made an announcement that $B(m, n)$ is *infinite* for $m \geqslant 2$ and $n \geqslant 72$. This, however, turned out to be premature. While the case of even $n$ remains unfounded, the gap in the argument for the case of odd $n$ took nearly nine years to patch. In their monumental 330-page, three-part paper, Novikov and Adian proved that $B(m, n)$ is indeed infinite for $m \geqslant 2$, $n \geqslant 4381$ and $n$ odd. This was in 1968, and, since then, the limit on $n$ has been lowered to $n \geqslant 665$, $n$ odd. Adian (1974) has expressed the possibility of lowering the limit to $n \geqslant 101$, and to $n \geqslant 33$ by some further alterations of their method. For $n < 33$, new techniques will be required and it appears that if $B(2,5)$ is, in fact, infinite then this fact may remain unresolved for a long time to come.

Britton (1973) has offered a shorter proof (280 pages) of the negative answer to the Burnside Problem. However, in his book, Adian (1975) has pointed out certain inconsistencies in Britton's argument which have since been acknowledged [see Britton (1983)]. Adian's book [see Adian (1979) for English translation] gives an improved version of the Novikov-Adian proof in a relatively compact form.

Ol'shanskii (1979) developed new geometric techniques and proved the existence of infinite nonabelian, *quasi-finite* groups [i.e., infinite groups all of whose proper subgroups are finite], solving an old problem of O. J. Schmidt [see Kurosh (1956) Vol. 1, page 57]. Encouraged by the Novikov-Adian solution of the Burnside problem, Ol'shanskii $(1982)_{\mathrm{a}}$ used these techniques and offered a significantly shorter proof of infiniteness of $B(m, n)$, $n$ odd, with a very large estimate for $n$, e.g., $n > 10^{10}$. To explain the strategy behind his proof, let $m \geqslant 2$ and $n \geqslant 2$ be fixed and let $F$ be an $m$-generator free group whose elements are totally ordered so that the words of shorter length precede those of larger length [see, for instance, Magnus et al., page 26]. We first construct an ascending chain (*) $N_1 < N_2 < \cdots$ of normal subgroups of $F$ as follows. Let $w_1$ be the least nontrivial element of $F$

and let $N_1 = \langle w_1^n \rangle^F$ be the normal closure of $w_1^n$ in $F$ [i.e., $N_1$ is the smallest normal subgroup of $F$ containing $w_1^n$]. Assume that we have already constructed $N_k = \langle w_1^n, \ldots, w_k^n \rangle^F$ for some $k \geqslant 1$. Let $w_{k+1}$ be an element of $F$ which is least with respect to the property that $w_{k+1} \notin N_k$ and modulo $N_k$, $w_{k+1}$ is of infinite order, and define $N_{k+1} = \langle w_1^n, \ldots, w_{k+1}^n \rangle^F$ to be the normal closure in $F$ of $N_k$ and $w_{k+1}^n$. Put $N = \bigcup_{k \geqslant 1} N_k$. Then clearly $F/N$ is periodic and the chain (*), when infinite, yields that $F/N$ is infinite. [Note that if the chain (*) is finite then we can not make any conclusion about finiteness of $F/N$]. Developing further the language of diagrams emerging from the geometric interpretation of defining relations due to van Kampen and Lyndon [see Lyndon and Schupp (1977), Chapter 5], Ol'shanskii $(1982)_{\mathrm{a}}$ proved that if $n$ is chosen to be odd and sufficiently large, e.g. $n > 10^{10}$, then it is indeed possible to construct an *infinite* chain (*) such that if $w_{k+1}$ is periodic modulo $N_k$ then it must be a conjugate, modulo $N_k$, of a power of $w_i$ for some $i \in \{1, \ldots, k\}$. From this it is immediate that each element of $F/N$ is of order dividing $n$ and hence $F/N \cong B(m, n)$ is infinite.

**2.** The General Burnside Problem was resolved by Golod in 1964. Based on his important joint work with Shafarevich, Golod proved the existence of finitely generated infinite $p$-groups for all primes $p$. Further examples, other than those of Novikov and Adian and Ol'shanskii, based on ideas from different areas of mathematics were produced by Aleshin (1972) [using automata theory], Sushchanskii (1979) [using permutation groups], Grigorchuk (1980) [using some ideas from functional analysis] and Gupta and Sidki (1983) [using automorphisms of trees]. Grigorchuk's example is a 3-generator, infinite 2-group and is particularly elegant [see Merzlyakov (1983) for a close relationship between Aleshin's and Grigorchuk's examples]. Here I give a construction, due to Gupta and Sidki, of a 2-generator, infinite $p$-group using tree automorphisms. First construct a 3-group.

Let $\mathbf{T}$ be a tree with root denoted "$\varnothing$" such that from $\varnothing$ and from the end of each branch precisely three new branches grow upwards [see FIGURE 1]. For any vertex $u$ of $\mathbf{T}$ let $\mathbf{T}(u)$ denote subtree with root "$u$." Then $\mathbf{T} = \mathbf{T}(\varnothing)$ is an infinite 3-regular tree with the property that $\mathbf{T}(\varnothing) \cong \mathbf{T}(u)$ for all $u$ [note that $\mathbf{T}(u) = (\mathbf{T}(u1), \mathbf{T}(u2), \mathbf{T}(u3))$, see FIGURE 2]. For each vertex $u$ we define an automorphism $t_u$ $(= [t_u : \mathbf{T}(u) \to \mathbf{T}(u)])$ of $\mathbf{T}(u)$ by mapping $\mathbf{T}(u1)$ onto $\mathbf{T}(u2)$, $\mathbf{T}(u2)$ onto $\mathbf{T}(u3)$ and $\mathbf{T}(u3)$ onto $\mathbf{T}(u1)$ [without shaking the trees]. Clearly $t_u$ has order 3. Using these automorphisms as a tool we next define an automorphism $a_u$ $(= a_u\colon \mathbf{T}(u) \to$

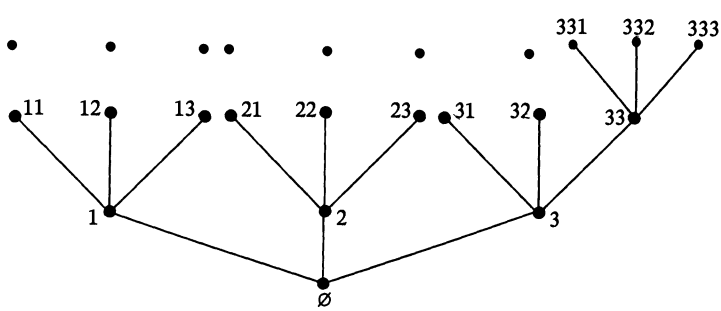



FIGURE 1.

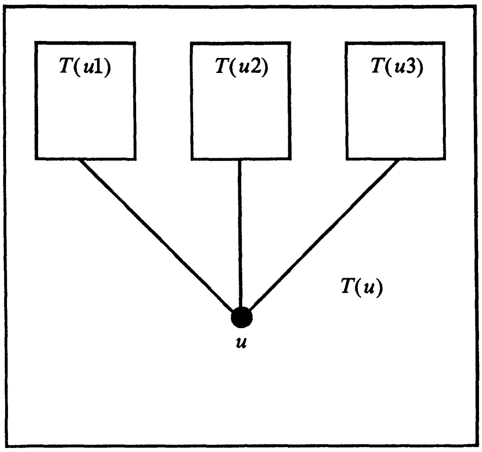



FIGURE 2.

$\mathbf{T}(u)$]) by its action on the subtrees $\mathbf{T}(u1)$, $\mathbf{T}(u2)$, and $\mathbf{T}(u3)$ as follows. We set $a_u = (t_{u1}, t_{u2}^2, a_{u3})$ with each component representing the action on the corresponding subtree. [This definition is unambiguous. Thus for instance, with $u = \varnothing$ we compute the image of $v = \varnothing 3321 = 3321$ under $a = a_\varnothing$ as $a_\varnothing(v) = a_3(3321) = a_{33}(3321) = t_{332}^2(3321) = 3323$.] Since $a_u^3 = (1, 1, a_{u3}^3)$, it follows by a simple induction on the length of the subscript label $u$ of $a$ that $a_u$ has order 3. We set $G_u = \langle a_u, t_u \rangle$ to be the subgroup of the automorphism group of $\mathbf{T}(u)$ generated by $a_u$ and $t_u$ [note that $G_\varnothing$ and $G_u$ are isomorphic for each vertex $u$]. We proceed to show that each $G_u$ so constructed is an infinite 3-group. We set $b_u = t_u^{-1} a_u t_u$ and $c_u = t_u^{-2} a_u t_u^2$. Then we have

$$a_u = \left(t_{u1}, t_{u2}^2, a_{u3}\right), \qquad b_u = \left(t_{u1}^2, a_{u2}, t_{u3}\right), \quad \text{and} \quad c_u = \left(a_{u1}, t_{u2}, t_{u3}^2\right)$$

[this can be verified directly]. Let $H_u = \langle a_u, b_u, c_u \rangle$. Then $H_u$ is a proper normal subgroup of $G_u$ of index 3. Since the action of $H_u$ on $\mathbf{T}(u1)$ is $G_{u1} \cong G_u$, $G_u$ must be infinite. To prove that $G_u$ is a 3-group, first observe that

$$G_u = \left\{ w_u t_u^{\alpha(u)}, \quad \text{where } w_u = w_u(a_u, b_u, c_u) \in H_u \quad \text{and } \alpha(u) \in \{0,1,2\} \right\}.$$

For each $g_u = w_u t_u^{\alpha(u)}$ we define its *formal length* $|g_u|$ as $\lambda(w_u)$ if $\alpha(u) \equiv 0 \bmod (3)$ and as $\lambda(w_u) + 1$ if $\alpha(u) \not\equiv 0 \bmod (3)$, where $\lambda(w_u)$ is the so-called *formal syllable length* of $w_u$ [e.g., $\lambda(a_u a_u b_u c_u a_u^{-1} c_u c_u) = 5$, whereas the total length is 8; see Magnus et al. p. 79 for the concept of syllable length]. To prove that $G_u$ is a 3-group it clearly suffices to prove by induction on the formal length $|g_u| = n$ that $g_u$ has order dividing $3^n$. If $|g_u| = 1$, then $g_u = a_u^\alpha$ or $b_u^\alpha$ or $c_u^\alpha$ or $t_u^\alpha$ for some $\alpha$, and it follows that $g_u$ has order 3. Let $|g_u| = n + 1$, $n \geqslant 1$, and assume the result for elements of formal length up to $n$. If $g_u$ is of the form $g_u = w(a_u, b_u, c_u)$, then, as an element of $H_u$, $g_u = (g_{u1}, g_{u2}, g_{u3})$, where each $g_{ui}$ has length at most $n$ or is of the form $g_{ui} = w_{ui} t_{ui}^{\alpha(i)}$ with $w_{ui} \in H_{ui}$, $|w_{ui}| \leqslant n$ and $\alpha(i) \in \{1,2\}$. Thus, without loss of generality, we may assume that $g_u$ itself has the form $g_u = w_u t_u^\alpha$ with $w_u \in H_u$, $|w_u| = n$ and $\alpha \in \{1,2\}$. Then $g_u^3$ is an element of $H_u$ which can be expressed in the form $g_u^3 = w_u(t_u^\alpha w_u t_u^{-\alpha})(t_u^{2\alpha} w_u t_u^{-2\alpha}) = (w_{u1}, w_{u2}, w_{u3})$, where each $w_{ui}$

is an element of $H_{ui}$ with $|w_{ui}| \leqslant n$. By the induction hypothesis it follows that $g_u^3$ has order dividing $3^n$ and, consequently, $g_u$ has order dividing $3^{n+1}$ as was to be proved. This completes our construction of 2-generator infinite 3-groups.

To construct $p$-groups for other odd primes, first consider a $p$-regular tree $\mathbf{T}(\varnothing)$, and define $t_u$: $\mathbf{T}(u) \to \mathbf{T}(u)$, for each vertex $u$, by cyclically permuting the subtrees $\mathbf{T}(u1), \ldots, \mathbf{T}(up)$, and, finally, define $a_u$: $\mathbf{T}(u) \to \mathbf{T}(u)$ by $a_u = (t_{u1}, t_{u2}^2, \ldots, t_{u(p-1)}^{p-1}, a_{up})$. For $p = 2$, consider the 4-regular tree $\mathbf{T}(\varnothing)$, define automorphisms $t_u$ as before and define $a_u = (t_{u1}, 1, t_{u3}^3, a_{u4})$. The remainder of the verifications are analogous to the case $p = 3$. We add that all these groups are known to have infinite exponent, i.e., the periodicity of these groups is unbounded.

**3.** *The order problem.* We recall Burnside's original question for the special case $n = 4$: *Is $B(m, 4)$ finite and if so, what is its order*? Sanov's solution of the finiteness of $B(m, 4)$ does very little towards the order problem of $B(m, 4)$. To see what is involved in this problem, we first note that, being a finite 2-group, $B(m, 4)$ is nilpotent of class precisely, say, $c(m)$. We abbreviate by writing $B(m)$ for $B(m, 4)$. It follows that if $w = w(x_1, \ldots, x_m)$ is an arbitrary element of $B(m)$, then $w$ can be uniquely expressed as the product $w = w_1 \ldots w_{c(m)}$, where $w_i \in \gamma_i(B(m))$ and $w_i \notin \gamma_{i+1}(B(m))$ [this can be achieved by a collecting process using repeatedly the identity $ab = ba[a, b]$]. If the order of the $i$th lower central factor $\gamma_i(B(m))/\gamma_{i+1}(B(m))$ is $2^{\alpha(i)}$, $i = 1, \ldots, c(m)$, then clearly $|B(m)| = 2^{\alpha(1) + \cdots + \alpha(c(m))}$. It is, therefore, natural to first ask a *weaker* version of the order problem, namely, the *Nilpotency Class Problem*: *what is the exact value of $c(m)$*?

Tobin (1954) proved that $c(2) = 5$ [and $|B(2)| = 2^{12}$]. For the general case, the most significant contribution to the class problem is due to Wright (1961), who developed a very high-power commutator calculus for $B(m)$ and proved that $c(m) \leqslant 3m - 1$. The problem acquired a renewed interest after Marshall Hall, Jr. [see Bruck (1963)] and Higman (1967) conjectured that *the free Burnside group of exponent* 4 *and countable infinite rank is solvable* [a group $G$ is said to be *solvable* if there exists a natural number $n$ such that $\delta_n(G) = 1$, where $\delta_1(G) = G$ and, for $i \geqslant 2$, $\delta_i(G) = \gamma_2(\delta_{i-1}(G))$]. A connection with the above conjecture of Hall and Higman and the class problem of $B(m)$ was first observed in Gupta and Gupta (1972). They proved that if, for some sufficiently large $m$, $c(m) < [5m/2]$, then $B(\infty)$ is indeed soluble. This was improved by Gupta and Quintana (1972) who proved that the same conclusion holds if, for some $m \geqslant 3$, $c(m) \leqslant 3m - 3$. For $m \geqslant 3$, Gupta and Newman (1974) improved Wright's result by proving that $c(m) \leqslant 3m - 2$. These last two results made the solubility conjecture most tantalizing. Bachmuth, Mochizuki and Weston (1973) provided evidence that $B(\infty)$ may not be solvable. This was finally confirmed by Razmyslov (1978), who developed some very sophisticated Lie ring techniques and constructed a nonsolvable group of exponent 4, thereby proving the nonsolvability of $B(\infty)$. Thus $c(m) = 3m - 2$ for $m \geqslant 3$.

The satisfactory solution to the class problem of $B(m)$ is only the first step towards its order problem. Finding the precise structure of the lower central quotients of $B(m)$ necessitates the use of computers. This has been successfully done to prove that $|B(3)| = 2^{69}$ [Bayes, Kautsky, and Wamsley (1974)] and $|B(4)| = 2^{422}$ [Alford, Havas and Newman (1975)]. For further progress on the order problem see Mann (1982).

For other applications of computers to questions like those of Burnside, the reader is referred to an interesting article by Havas and Newman (1980) (see also Newman (1984) for a computer-based proof of finiteness of $B(m,6)$). An attempt has been made by Grunewald, Havas, Mennicke, and Newman (1980) to study the finiteness of the Burnside group $B(2,8)$. This work occupies 160 pages of calculations studying certain subgroups of $B(2,8)$. Even the restricted Burnside problem for $B(2,8)$ remains unresolved. For the current status of the restricted Burnside problem see the very informative survey article by Vaughan-Lee (1985).

**4.** *Some Open Problems.* As has been mentioned, the methods of Novikov and Adian and those of Ol'Shanskii remain unsuitable for the Burnside problem for small exponents. Here I list some of the open problems which are of immediate interest. I begin by asking:

**Problem 1.** *Is $B(2,5)$ finite?*

Apart from exponents 2 and 4, the Burnside problem for other 2-power exponents remains wide open. Here we record a weaker version for $B(2,8)$:

**Problem 2.** *Solve the restricted Burnside problem for $B(2,8)$.*

Kostrikin's solution of the restricted Burnside problem is valid for groups of prime exponent and we are tempted to record the smallest case for $B(m, p^k)$:

**Problem 3.** *Solve the restricted Burnside problem for $B(2,9)$.*

A finitely generated group is said to be *finitely presented* if it can be defined by finitely many relations (see Magnus et al., page 7). The infinite groups of Novikov and Adian, Ol'Shanskii, Grigorchuk, and Gupta and Sidki are all known to be infinitely presented. In fact the verification of the periodicity requires these groups to have infinitely many relations. It is, therefore, natural to reverse the situation and ask whether or not a periodic group becomes finite if its periodicity is the consequence of finitely many relations. In other words: *Are finitely presented periodic groups finite?* (see Lyndon (1984), Problem 10b). The following weaker version of this problem remains wide open even for exponent 5:

**Problem 4.** *Are finitely presented groups of prime exponent finite?*

Finally, analogous to the formulation of the restricted Burnside problem, we formulate:

**Problem 5.** *Is there a bound on the orders of finite quotients of finitely presented groups of prime-power exponent?*

**5.** *Some Applications.* This section has nothing to do with the Burnside Problem. I include it just to point out how the techniques developed to answer the Burnside's questions have revealed and continue to reveal some of the very deep facts in group theory.

In 1968 Adian proved that for all odd $n \geqslant 4381$, there is an algorithm to decide whether or not an arbitrary element of $B(m, n)$ is the identity element. Using the recursive presentation of $B(m, n)$ so developed, he further proved in 1971 that for each nontrivial element $x \in B(m, n)$, there exists an element $z \in B(m, n)$ such that $x \in \langle z \rangle$, and whenever $y$ commutes with $x$ then $y \in \langle z \rangle$. It follows immediately from this that the abelian subgroups of $B(m, n)$ are cyclic and that the centre

of $B(m, n)$ is trivial for odd $n \geqslant 4381$. Further analysis yields: *Every finite subgroup of* $B(m, n)$, $m \geqslant 2$, *n odd*, $n \geqslant 4381$, *is cyclic* [see Adian 1975].

The additive group of rationals is an example of a torsion-free group which has the property that every two nontrivial elements are powers of a single element. The existence of nonabelian torsion-free groups with the above property was a long-standing problem. In 1971 Adian developed a certain system $\mathbf{M} = \mathbf{M}(a_1, \ldots, a_m)$ of *elementary words* and introduced, for each odd $n \geqslant 4381$, a group $A(m, n) = \langle a_1, \ldots, a_m, d | a_i d = da_i \ (i = 1, \ldots, m), a^n = d$ for all $a^n \in \mathbf{M} \rangle$. It turns out that the group $A(m, n)$ is torsion-free and that $B(m, n)$ is the factor group of $A(m, n)$ by its cyclic centre $\langle d \rangle$, yielding the much stronger result: *For each odd* $n \geqslant 4381$, *there exists a finitely generated torsion-free group with cyclic centre satisfying the identical relation* $[x^n, y] = 1$ [see Adian (1975)]. Another interesting result proved by Adian in 1971 was the fact that, for large odd $n$, the group $B(3, n)$ is embedded in $B(2, n)$. This has since been improved to: *For n odd*, $n \geqslant 4381$, $B(m, n)$ *is embedded in* $B(2, n)$ *for all* $m \geqslant 3$ [Shirvanian (1976); see Adian (1975) for the case $m = 3$].

The geometric methods developed by Ol'shanskii are quite powerful and effective. These have been successfully used to resolve many outstanding problems in group theory. Among the results proved using his techniques are: *There exists an infinite simple group all of whose proper subgroups are infinite cyclic* [Ol'shanskii (1979)]; *For every prime* $p > 10^{75}$, *there exists an infinite simple p-group all of whose proper subgroups are of order p* [Ol'shanskii $(1982)_b$]; *There exists a nonabelian variety of groups all of whose finite groups are abelian* [Ol'shanskii (1985)]; *There exist finitely generated nonabelian divisible groups* [Guba (1986)]; *There exists a 2-generator group G with an element* $a \in G$ *such that every element of G is conjugate to some power of a* [Guba (1986)].

In Problem #5603, *Amer. Math. Monthly*, 75 (1986) 685–686, John Milnor asked if there exists a finitely generated group whose growth function (the one which measures the number of distinct elements of a given length) is neither polynomial nor exponential. The evidence was much against the existence and the question has since generated considerable interest in the literature. The infinite 2-groups constructed by Grigorchuk turned out to have subexponential growth: *There exists a finitely generated (periodic) group whose growth function is neither polynomial nor exponential* [Grigorchuk (1983)]. [In a subsequent paper, Grigorchuk further proves that the tree groups constructed in Section 2 also have subexponential growth.]

It is an elementary fact that every finite $p$-group is embedded in an $n$-fold wreath tower $(((C_p wr C_p) \cdots) wr C_p)$ of cyclic groups of order $p$ [see, for instance, Hall (1959)]. The infinite $p$-groups constructed by Gupta and Sidki using the tree automorphisms have the additional property that they contain these $n$-fold wreath towers for all $n$. Further, for $p$ odd, these groups have the property that the normal closure of any nontrivial element contains some term of the derived series. We thus have: *For every prime p, there exists a 2-generator infinite p-group which embeds every finite p-group as a subgroup and in which, for p odd, every proper quotient is finite* [Gupta and Sidki (1983)]. Finally, these tree constructions further yield: *For every* $n > 2$, *there exists a finitely generated p-group G with a subgroup H of finite index such that* $H \cong G^n$, *the direct product of n copies of G* [Gupta and Sidki (1984)].

M. F. Newman (1980) has compiled a comprehensive bibliography of papers up to 1979 which are connected with the Burnside problem. A count reveals that it has over 225 items. There have been dozens more since then. One has to agree with Chandler and Magnus (1982) that, very much like "Fermat's last theorem" in number theory, Burnside's problem has acted as a catalyst for research in group theory. The fascination exerted by a problem with an extremely simple formulation which then turns out to be extremely difficult has something irresistible about it to the mind of the mathematician.

**Acknowledgement.** I thank Yuri Bahturin for many useful conversations during the preparation of this article and for supplying some of the latest references. I also thank M. F. Newman and Peter M. Neumann for their helpful comments on an earlier draft.

REFERENCES

S. I. Adian (1974) Periodic groups of odd exponent, *Lecture Notes in Math*, 372, Springer-Verlag, 8–12.

______, (1975), The Burnside problem and identities in groups, *Grenzgeb Math.* 95, Springer-Verlag 1979.

______, (1984), Investigation of the Burnside problem and related questions, *AN. SSR*, 168, (Trudy Mat. Inst.) 171–196. [Russian]

S. I. Adian and N. N. Repin (1986), Exponentiality of the lower bound for the nilpotency class of Engel Lie algebras, *Mat. Zametki*, 39, 444–452. [Russian]

S. V. Aleshin (1972), Finite automata and the Burnside problem for periodic groups, *Mat. Zametki*, 11, 319–328. [Engl. Transl. *Math. Notes*, 11, 199–203]

William A. Alford, George Havas, and M. F. Newman (1975), Groups of exponent four, *Notices Amer. Math. Soc.*, 22, A.301.

S. E. Arshon (1937), Proof of the existence of $n$-valued infinite asymmetric sequences, *Mat. Sbornik*, 2(44), 769–779. [Russian]

S. Bachmuth, H. Y. Mochizuki, and K. Weston (1973), A group of exponent 4 with derived length at least 4, *Proc. Amer. Math. Soc.*, 39, 228–234.

A. J. Bayes, J. Kautsky, and J. W. Wamsley (1974), Computations in nilpotent groups (applications), *Proc. Second Internat. Conf. Theory of Groups*, Canberra, Lecture Notes in Math. 372 (Springer-Verlag) 82–89.

J. L. Britton (1973), The existence of infinite Burnside groups, Word Problems, North Holland, pp. 67–348.

______, (1980), Erratum: The existence of infinite Burnside groups, Word Problems II, The Oxford Book, North Holland, p. 71.

R. H. Bruck (1963), Engel conditions in groups and related questions, *Lecture Notes* (Canberra).

W. Burnside (1902), On an unsettled question in the theory of discontinuous groups, *Quart. J. Pure Appl. Math.*, 33, 230–238.

W. Burnside (1905), On criteria of finiteness of the order of a group of linear substitutions, *Proc. London Math. Soc.*, (2) 3, 230–238.

Richard A. Dean (1965), A sequence without repeats in $x, x^{-1}, y, y^{-1}$, *Amer. Math. Monthly* 72, 383–385.

E. S. Golod (1964), On nil-algebras and finitely approximable $p$-groups, *Izv. Akad. Nauk. USSR Ser. Mat.*, 28, 273–276. [Russian]

R. I. Grigorchuk (1980), On the Burnside problem for periodic groups, *Funkcional Anal. i Prilozen*, 14, 53–54 [Russian]. Engl. transl. *Functional Anal. Appl.*, 14, 41–43.

______, (1983), On Milnor's problem of group growth, *Soviet Math. Dokl.*, 28, 23–26.

Otto Grün (1940), Zusammenhang zwischen Potenzbildung und Kommutatorbildung, *J. reine angew. Math.*, 182, 158–177.

Fritz J. Grunewald, George Havas, J. L. Mennicke and M. F. Newman (1980), Groups of exponent eight, Proc. Workshop on Burnside Groups, Bielefeld (1977), *Springer Lecture Notes*, 806, 49–188.

V. S. Guba (1986), A finitely generated divisible group, *Izvestija Akad. Nauk SSSR Ser. Mat.*, 50, 883–924.

C. K. Gupta and N. D. Gupta (1972), On groups of exponent four II, *Proc. Amer. Math. Soc.*, 31, 360–362.

N. D. Gupta and M. F. Newman (1974), The nilpotency class of finitely generated groups of exponent four, *Proc. Second Internat. Conf. Theory of Groups*, Canberra, Springer Lecture Notes 372, 330–332.

N. D. Gupta and R. B. Quintana, Jr. (1972), On groups of exponent four III, *Proc. Amer. Math. Soc.*, 33, 15–19.

Narain Gupta and Said Sidki $(1983)_a$, On the Burnside problem for periodic groups, *Math. Zeitschr.*, 182, 385–388.

______, (1984), Extension of groups by tree automorphisms, *Contemporary Math.*, 33, 232–246.

______, $(1983)_b$, Some infinite $p$-groups, *Algebra i Logika*, 22, 584–589. [Russian Edition]

Marshall Hall Jr. (1958), Solution of the Burnside problem for exponent sixsMath, *Ill. J. Math.*, 2, 764–786.

______, (1959), The Theory of Groups, Macmillan, New York.

______ and Charles Sims (1982), The Burnside group of exponent 5 with two generators, *Proc. Group Theory Conf.*, St. Andrews (1981), *London Math. Soc. Lecture Notes*, 71, 207–220.

P. Hall and Graham Higman (1956), On the $p$-length of $p$-soluble groups and reduction theorems for Burnside's problem, *Proc. London Math. Soc.*, (3) 6, 1–42.

George Havas and M. F. Newman (1980), Applications of computers to questions like those of Burnside, *Proc. Workshop on Burnside Groups*, Bielefeld (1977), *Lecture Notes in Math.*, 806 (Springer-Verlag) 211–230.

George Havas, G. E. Wall and J. W. Wamsley (1974), The two generator restricted Burnside group of exponent five, *Bull. Austral. Math. Soc.*, 10, 459–470.

Graham Higman (1956), On finite groups of exponent five, *Proc. Cambridge Phil Soc.*, 54, 1–4.

______, (1967), The orders of relatively free groups, *Proc. Internat. Conf. Theory of Groups*, Canberra (1965) Gordon & Breach, N.Y., 153–165.

A. I. Kostrikin (1955), Solution of a weakened problem of Burnside for exponent 5, *Izv. Akad. Nauk SSR Ser. Mat.*, 19, 233–244. [Russian]

______, (1959), The Burnside problem, *Izv. Akad. Nauk. SSSR Ser. Mat.*, 23, 3–34. [Russian] [*Amer. Math. Soc. Transl.*, (2) 36 (1964), 63–99.]

______, (1979), Sandwiches in Lie algebras, *Mat. Sbornik*, 110, 3–12. [Russian]

______, (1986), Around Burnside, *Nauka*, *Moscow*. [Russian]

A. G. Kurosh (1955), The Theory of Groups, Chelsea Publishing, N.Y.

Friedrich Levi and B. L. van der Waerden (1933), Über eine besondere Klasse von Gruppen, *Abh. Math. Sem.*, 9, 154–158, Univ. Hamburg.

R. C. Lyndon (1954), On Burnside's problem, *Trans. Amer. Math. Soc.*, 77, 202–215.

______, (1955), On Burnside's problem II, *Trans. Amer. Math. Soc.*, 78, 329–332.

______, (1984), Problems in combinatorial group theory, *Proc. Utah Conf.*, *Annals of Math Studies* III (1987), 2–33, Princeton University Press.

Roger C. Lyndon and Paul E. Schupp (1977), Combinatorial Group Theory, *Ergeb. Math.*, 89, Springer-Verlag, N.Y.

Wilhelm Magnus (1950), A connection between the Baker-Hausdorff formula and a problem of Burnside, *Ann. of Math.*, (2) 52, 111–126.

Wilhelm Magnus, Abraham Karrass, and Donald Solitar (1966), Combinatorial Group Theory, Interscience, N.Y.

A. J. S. Mann (1982), On orders of groups of exponent four, *J. London Math. Soc.*, (2) 26, 64–76.

Yu. I. Merzlyakov (1983), Infinite finitely generated periodic groups, *Dokl. Akad. Nauk SSSR*, 268, 803–805. [Russian]

M. Morse and C. A. Hedlund (1944), Unending chess, symbolic dynamics and a problem in semigroups, *Duke Math. J.*, 11, 1–7.

M. F. Newman (1984), Groups of exponent six, Computation Group Theory (Durham, 1982) 39–41. [Academic Press, London, New York]

P. S. Novikov (1959), Solution of Burnside's problem on periodic groups, *Uspehi Mat. Nauk*, (N.S.), 14 (89) 236–237. [Russian]

P. S. Novikov and S. I. Adian (1968), Infinite periodic groups I, II, III, *Izv. Akad. Nauk SSSR Ser. Mat.*, 32, 212–244, 251–254, 709–731. [Russian]

A. Ju. Ol'shanskii (1979), Infinite groups with cyclic subgroups, *Dokl. Akad. Nauk SSSR*, 245, 785–787. [Russian]

______, $(1982)_a$, On the Novikov-Adian theorem, *Math. USSR Sbornik*, 46, 203–236. [Russian]

______, $(1982)_b$, Groups of bounded period with subgroups of prime order, *Algebra i Logika*, 21, 553–618.

A. Ju. Ol'shanskii (1985), Varieties in which all finite groups are abelian, *Math. Sbornik*, 126 (168), 59–82 [Russian] [English transl., *Math. SSSR Sbornik*, 54 (1986), 57–80]

Ju. P. Razmyslov (1978), On a problem of Hall-Higman, *Izv. Akad. Nauk SSSR Ser. Mat.*, 42, 833–847. [Russian]

I. N. Sanov (1940), Solution of Burnside's problem for exponent four, Leningrad Gos. Univ. Ped. Inst. *Uc. Zap. Mat. Ser.*, 10, 166–170. [Russian]

V. L. Shirvanian (1976), Embedding of the group $B(\infty, n)$ in the group $B(2, n)$, *Izv. Akad. Nauk SSSR Ser. Mat.*, 40, 190–208. [Russian]

V. I. Sushchanskii (1979), Periodic $p$-groups of permutations and the unrestricted Burnside problem, *Dokl. Akad. Nauk SSSR*, 247, 557–561. [Russian]

John Joseph Tobin (1954), On groups with exponent 4, Ph.D. thesis, Univ. of Manchester.

M. R. Vaughan-Lee (1985), The restricted Burnside problem, *Bull. London Math. Soc.*, 17, 113–133.

C. R. B. Wright (1961), On the nilpotency class of a group of exponent four, *Pacific J. Math.*, 11, 387–394.

**Department of Ruined Reputations**

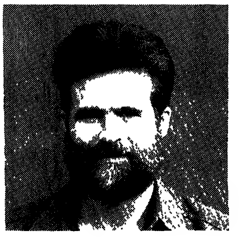
The John von Neumann National Supercomputer Center is owned and operated by the Consortium for Scientific Computing, compromising the University of Arizona, Brown, the University of Colorado, Columbia, Harvard, the Institute for Advanced Study, MIT, New York University, Pennsylvania State University, the University of Pennsylvania, Princeton, the University of Rochester, and Rutgers.

—from Announcement of Symposium on Automatic Groups

# The Shortest Planar Arc of Width 1*

ANI ADHIKARI, *Stanford University*
JIM PITMAN, *University of California, Berkeley*

ANI ADHIKARI grew up in Calcutta, where she obtained a B.S. at the Indian Statistical Institute. She obtained a Ph.D. in statistics at the University of California, Berkeley (1986), and is now an Assistant Professor at the Department of Statistics, Stanford University. She is interested in probability theory, applied statistics, and statistical computing. She enjoys PostScript graphics, a sample of which can be found in this paper.



JIM PITMAN grew up in Hobart, Tasmania. After obtaining a B.S. at the Australian National University, Canberra, he obtained a Ph.D. in probability theory at the University of Sheffield (1974). Since then he has mostly been at the University of California, Berkeley, where he is now a Professor in the Department of Statistics. He spent two years at Cambridge University and shorter periods in Copenhagen and Paris. His main research interest is Brownian motion and related stochastic processes. He is currently completing an undergraduate textbook on probability.

**1. Introduction.** A floor is ruled with parallel lines spaced unit distance apart. You are given a piece of wire of length $l$ which you are free to bend but not stretch. Can you bend the wire so that if dropped on the floor the bent piece of wire is certain to cross at least one of the lines, no matter how it falls? In this article we find the least $l$ such that this can be done, and show how the wire should be bent.

More formally, let $\alpha$: $[0,1] \to \mathbf{R}^2$ be a continuous rectifiable arc, of length $l(\alpha)$. For $0 \leqslant \theta \leqslant \pi$, define the *width of $\alpha$ between parallels at angle $\theta$* by

$w_\theta(\alpha)$ = distance between supporting parallel lines at an angle $\theta$ to the $x$-axis.

And define the *width of $\alpha$* by

$$w(\alpha) = \inf_{0 \leqslant \theta \leqslant \pi} w_\theta(\alpha).$$

Our problem is to identify an arc of minimal length among all arcs of width at least 1. A circular arc of diameter 1 has length $\pi = 3.14\ldots$. Three sides of the unit square reduces the length to 3. Two sides of an equilateral triangle of altitude 1 further reduces the length to $4/\sqrt{3} = 2.309401\ldots$. The minimal length turns out only slightly less, namely $2.27829\ldots$, and is achieved by bending the wire into a shape which we call the *caliper*, shown in FIGURE 2.

*Research supported in part by NSF grant DMS-8502930.

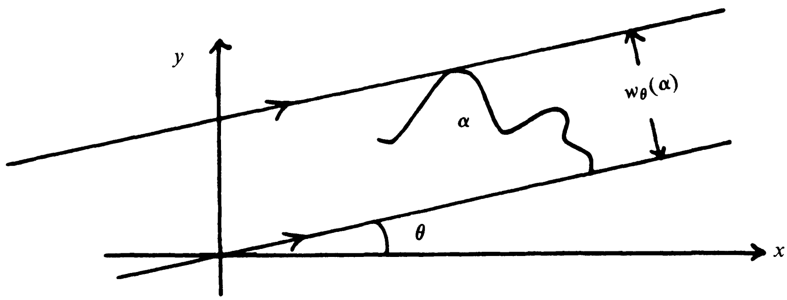



FIG. 1. The width between parallels at angle $\theta$.

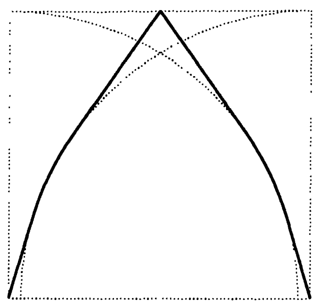


FIG. 2. The caliper.

In Section 2 we identify the caliper of length 2.27829... as the shortest convex arc of width 1. In Sections 3 and 4 we show that the caliper is in fact shortest among all arcs, convex and nonconvex. For the most part, our arguments are geometric, in the spirit of Kazarinoff [3], Yaglom and Boltyanskii [5], and Niven [4].

Our problem is a deterministic variation of Buffon's famous needle problem: in case you do not bend the wire at all, but leave it as a straight needle of length $l$, what is the chance that the needle crosses at least one line if dropped at random on the floor? As observed by Bertrand in the last century, and argued in detail by Barbier [1], under natural assumptions of randomness the expected number of crossings of the grid by a needle bent into a planar curve is a constant $c$ times the length $l$, regardless of the shape of the curve. A closed convex curve of constant width 1 must cross the lines exactly twice, no matter how it falls. Since a circle of diameter 1 has constant width 1 and length $\pi$, it follows that

(i) the constant is $c = 2/\pi$,

(ii) every closed convex curve of constant width 1 has the same length $\pi$, and

(iii) for $l \leqslant 1$ the probability that a randomly dropped straight needle crosses the lines is $(2/\pi)l$.

For other questions about randomly dropped curves, and further references, see DeTemple and Robertson [2].

The problem solved in this article is a special case of a stochastic geometry problem which we do not know how to solve for all $l$. Given a wire of length $l$, how should it be bent to maximize the probability that it crosses at least one line when tossed at random on the ruled floor? Here we just find the least length $L$ such that

this maximum probability is one. For $0 < l \leqslant 1$ the best strategy is easily shown to be leaving the wire straight, with crossing probability $(2/\pi)l$. But for $1 < l < L$ we do not know how to bend the wire to maximize the crossing probability. Some remarks on this and other problems in the same vein are mentioned in the final section.

*Notation.* Since rotation, translation, and reflection of an arc in the plane affect neither its length nor its width, we will often use these operations to reduce calculations to arcs with some convenient orientation. A generic arc $\alpha$ will often be denoted instead $A^{\sim}B$ to indicate that it starts at $A$ and ends at $B$. To indicate the arc is convex we write $A^{\wedge}B$ instead of $A^{\sim}B$. For points $A$ and $B$ in the plane we use the following notation:

$\overline{AB}$ for the straight line through $A$ and $B$,

$AB$ for the segment of this line between $A$ and $B$,

$|AB|$ for the length of this line segment.

Constructions of arcs may be indicated by notation such as this: if $A^{\sim}B$ is some arc under consideration, then $A^{\sim}BA$ is the closed arc obtained by following $A^{\sim}B$ from $A$ to $B$, then returning to $A$ along a line segment. In such constructions the precise parameterization of the arc will be irrelevant, though the order in which points are visited may be important.

**2. The Shortest Convex Arc of Width One.** Suppose now that the arc $A^{\wedge}B$ is *convex*, meaning that the closed arc $A^{\wedge}BA$ is the boundary of a convex subset of the plane:

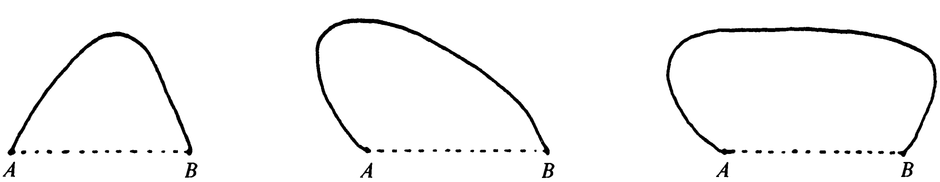



FIG. 3. Some convex arcs.

A convex arc $A^{\wedge}B$ must lie on one side of the line $\overline{AB}$. Without loss of generality, we take $\overline{AB}$ to be horizontal, and suppose $A^{\wedge}B$ lies above $\overline{AB}$. We propose to find the shortest such arc of width one. In case $A^{\wedge}B$ extends to the left of $A$ or to the right of $B$, a shorter arc with greater width is obtained by dropping a perpendicular to $\overline{AB}$:

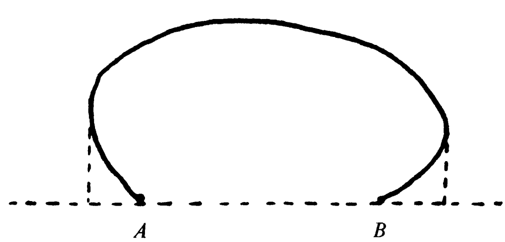



FIG. 4. Reduction to arcs over an interval.

So we may assume that $d = |AB| \geqslant 1$, and confine our search to arcs $A\hat{}B$ lying in the strip above $AB$, such as the leftmost arc in FIGURE 3. In this case the width condition can be reformulated more simply as follows, with reference to the horizontal segment $A'B'$ at unit height above $AB$, and open discs $Disc(A)$ and $Disc(B)$ centered at $A$ and $B$, with radii 1, as shown in the next figure:

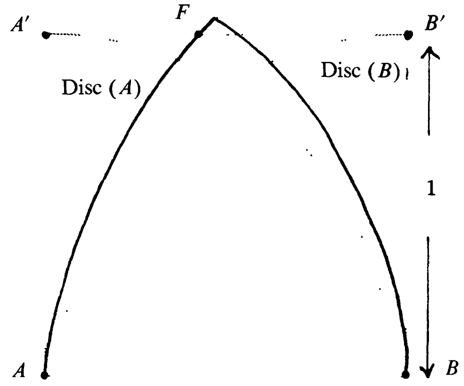



FIG. 5. An arc of width greater than 1.

LEMMA 1. *For a convex arc* $A\hat{}B$ *above* $AB$ *with* $|AB| \geqslant 1$, *width* $(A\hat{}B) \geqslant 1$ *if and only if the following three conditions all hold*:
(i) $A\hat{}B$ *intersects* $A'B'$ *at some point* $F$.
(ii) $A\hat{}F$ *does not intersect* $Disc(B)$.
(iii) $F\hat{}B$ *does not intersect* $Disc(A)$.

*Proof of Sufficiency.* Suppose the three conditions hold. Width at least 1 in the horizontal and vertical directions is clear. For the width between down-sloping parallels, consider two parallels distance 1 apart, one through $A$, the other above it tangent to $Disc(A)$. The upper parallel must hit the arc between $F$ and $B$ by (i) and (iii). Similarly, width at least 1 between up-sloping parallels is implied by (i) and (ii).

*Proof of Necessity.* If (i) fails, the vertical width is less than 1. The only remaining case is if (i) holds but either (ii) or (iii) fails. If say (ii) fails, there exist points $C$ and $D$ on $A\hat{}B$ as in the following diagram such that $C$ and $D$ lie on $Circle(B)$ bounding $Disc(B)$, and apart from these points arc $C\hat{}D$ lies inside the open $Disc(B)$.

By convexity of $A\hat{}B$, the arc lies inside the shaded region in the diagram, with boundary defined in terms of $\overline{CD}$ and $Disc(B)$. The upper supporting line of $A\hat{}B$ parallel to $CD$ must, therefore, intersect $Disc(B)$, giving a width less than 1. $\square$

According to the lemma, for fixed $A$ and $B$ at distance $d \geqslant 1$, the shortest convex arc $A\hat{}B$ of width 1 lying above $AB$ is the shortest path from $A$ to $B$ via some $F$ on $A'B'$, not entering $Disc(B)$ on the way from $A$ to $F$, and not entering $Disc(A)$ on the way from $F$ to $B$.

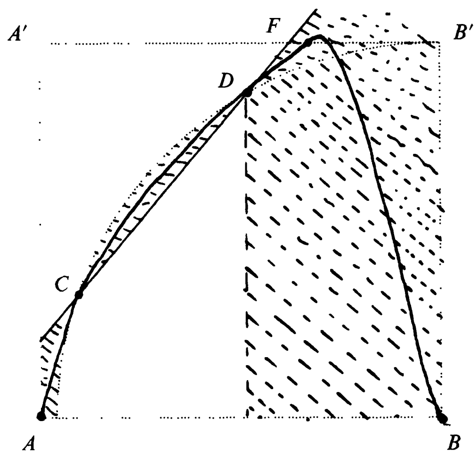



FIG. 6. An arc satisfying (i) but not (ii).

In case $d \geqslant 2/\sqrt{3}$, this shortest path is obvious. For $F$ the midpoint of $A'B'$ the straight lines $AF$ and $FB$ do not meet the discs, so $AFB$ is the shortest, with length $\sqrt{4+d^2}$. For $d = 2/\sqrt{3}$, the length is 2.309401... .

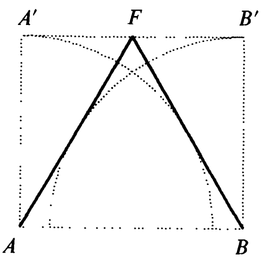



FIG. 7. Two sides of an equilateral triangle: $d = 2/\sqrt{3}$.

If $d < 2/\sqrt{3}$, the constraint of the discs is felt. For any $F$ on $A'B'$, it is easy to see that the shortest way to get from $A$ to $F$ without entering $Disc(B)$ is to go along

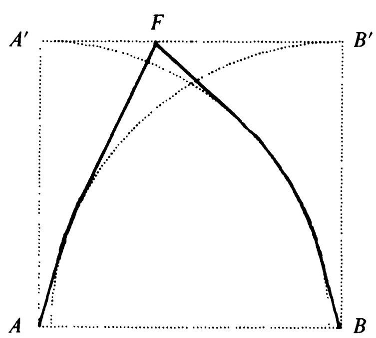



FIG. 8. The shortest path from $A$ to $B$ via $F$ satisfying (i), (ii), (iii).

$AF$, if this can be done without entering $Disc(B)$; else to go from $A$ to $Circle(B)$ along a tangent, stay on the circle awhile, and leave for $F$ along another tangent. And the shortest way to return from $F$ to $B$ without entering $Disc(A)$ can be described similarly, with $Circle(A)$ replacing $Circle(B)$.

To find the shortest convex arc from $A$ to $B$ satisfying (i)–(iii), reflect the arc in $A'B'$ after it first touches $A'B'$ at $F$. The problem reduces to finding the shortest arc from $A$ to $B^*$, never entering $Disc(B)$ or $Disc(A^*)$, as in FIG. 9.

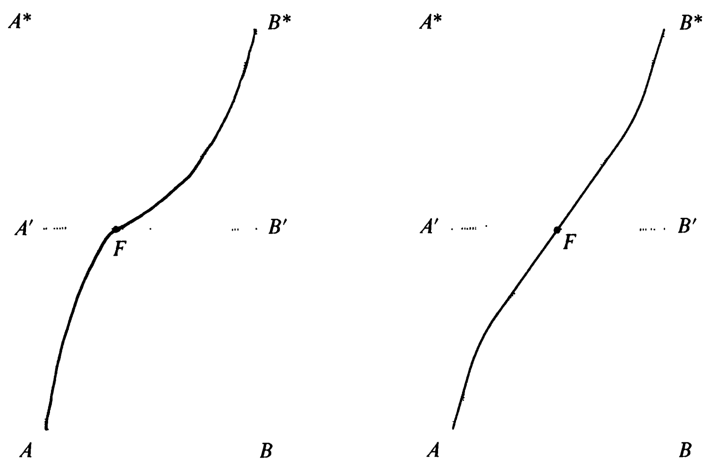



FIG. 9. Paths from $A$ to $B^*$ avoiding the discs.

It is easy to see that the minimal arc must be straight as it passes through $F$. This can only happen if $F$ is the midpoint of $A'B'$, so that the minimal arc is symmetric.

By symmetry, the length of this shortest arc is $2l(d)$ where $l(d)$ is the length of the arc between $A$ and $F$, $d = |AB|$, and $1 \leqslant d \leqslant 2/\sqrt{3}$. By the elementary geometry of FIGURE 10,

$$l(d) = \sqrt{d^2 - 1} + \left[\pi/2 - \arctan\sqrt{d^2 - 1} - 2\arctan(d/2)\right] + d/2.$$

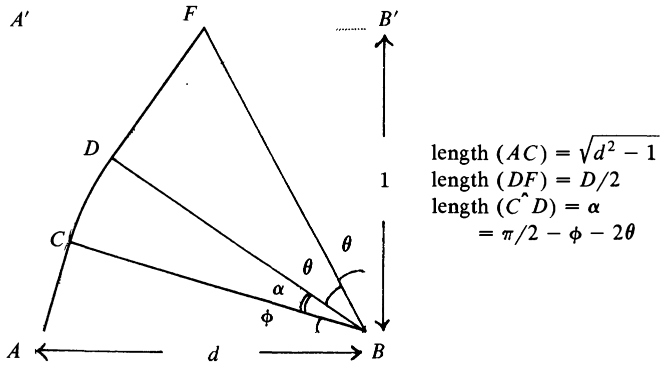



FIG. 10. Calculation of $l(d)$.

To find the $d$ which minimizes this length, differentiate $l$ with respect to $d$:

$$l'(d) = \frac{d}{\sqrt{d^2-1}} + \left[ -\frac{d}{\sqrt{d^2-1}} \cdot \frac{1}{1+(d^2-1)} - \frac{1}{1+d^2/4} \right] + 1/2$$

$$= \frac{\sqrt{d^2-1}}{d} + \frac{1}{2}\left( \frac{d^2-4}{d^2+4} \right).$$

Set this equal to 0 to see that the minimum is achieved when $z = d^2$ satisfies

$$4(z-1)(4+z)^2 - z(z-4)^2 = 0.$$

Solving the cubic yields the minimizing $d^* = 1.04359\ldots$ and $l(d^*) = 2.27829\ldots$. The shortest convex arc of width 1 so obtained is shown in the center of the next figure. We call this arc the *caliper*.

Also shown are the arcs considered above for other values of $d$ near $d^*$. For $d \geqslant d^*$ the argument above shows these are the shortest convex arcs of width 1 with endpoints distance $d$ apart. But for $d \leqslant d^*$ we are unable to draw this conclusion due to the initial step of dropping perpendiculars to reduce to the case of arcs over the interval between the ends. All we can say is that these arcs are shortest among arcs of width 1 that stay in the strip over the interval.

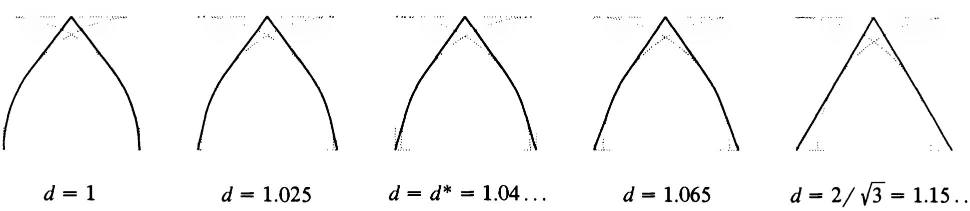



FIG. 11. The caliper for various values of $d$.

**3. Results for Nonconvex Arcs.** The aim of this section and the next is to show that the caliper defined in Section 2 is in fact a shortest arc of width 1 among all arcs, not just among convex ones as shown already. This seems to be a lot harder than you might expect. Intuitively, it is hard to believe that you could do any better by allowing the arc to cross through its convex hull. But a difficulty can be seen from the existence of arcs such as those shown below for which there is no shorter arc with the same convex hull.

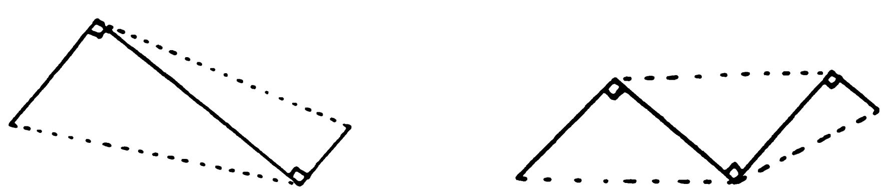


FIG. 12. Arcs of minimal length with given convex hull.

It must somehow be argued that such arcs are still far from the shortest for the given width. A first step in this direction is provided by the following inequality:

PROPOSITION 1. *Let $l$ be the length of an arc, $d$ the distance between its ends, and $w$ its width. Then $l \geqslant \sqrt{4w^2 + d^2}$.*

*Proof.* By definition of $w$, in every direction there are two parallel lines distance $w$ apart, each touched by the arc. Consider two such parallels, taken parallel also to the lines between the endpoints in case $d > 0$, as in the left panel below. Reflecting a portion of the arc as in the right panel produces an arc of the same length $l$ joining points at a distance of $\sqrt{4w^2 + d^2}$ by Pythagoras. □

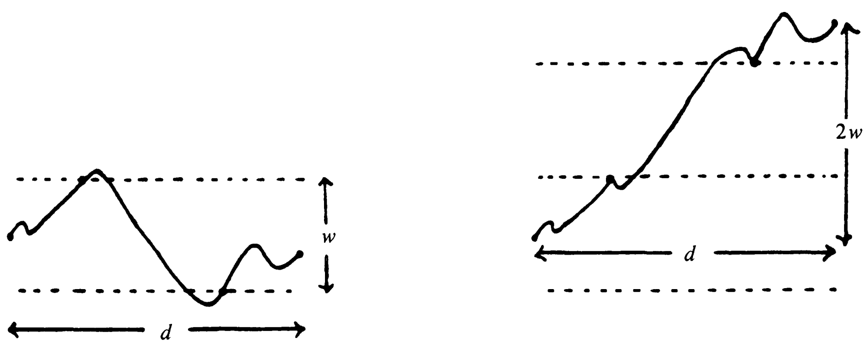



FIG. 13. Construction for the proof of Proposition 1.

*Remarks.* For width $w = 1$ all we get immediately is $l \geqslant 2$, using $d \geqslant 0$. But this is very crude. For $d \geqslant 2/\sqrt{3}$ the above bound is attained by two line segments, as remarked already in Section 2. For $0 \leqslant d < 2/\sqrt{3}$ it is not hard to show that the bound cannot be attained. In the case of a closed convex curve, when $d = 0$, the least possible length is $\pi$, attained by any curve of constant width 1. This result is given for closed curves in Exercise 7-18 of Yaglom and Boltyanskii [5]. A quick proof in this case can be given by consideration of the expected number of crossings in the Buffon setting described in the introduction. This result can be extended to nonconvex closed curves by an easy variation of arguments in this section. But we do not know the best inequality for $0 < d < 2/\sqrt{3}$.

To get a handle on possibly nonconvex arcs, we work with the class $\mathbf{C}_M$ of all arcs starting at a fixed point $A$ of length at most 3, consisting of at most $M$ line segments, where $M$ is a fixed integer. It is easy to show that length and width are continuous functions for a topology which makes this class of arcs compact. This class therefore contains a shortest arc of width at least 1. Pick such a *minimal arc* and call it $\alpha^*$. We propose to study the properties of $\alpha^*$, as a preliminary to getting lower bounds on its length which do not involve $M$, the bound on the number of segments. Clearly, any such lower bound is also a lower bound on the length of any arc whatsoever with width one.

LEMMA 2. *The sequence of endpoints of segments of a minimal arc $\alpha^*$ is a permutation of the set of corners of the convex hull of $\alpha^*$.*

*Proof.* Let $H$ be the set of corners of the convex hull of $\alpha^*$. The proof is in two steps.

(i) Each element of $H$ is visited exactly once by $\alpha^*$.

(ii) Each endpoint of a segment of $\alpha^*$ is an element of $H$.

*Proof of* (i). Clearly every corner of the convex hull is an endpoint of some line segment of $\alpha^*$, so visited at least once by $\alpha^*$. Suppose some element $h \in H$ is visited more than once by $\alpha^*$. Then an arc with one less segment formed by cutting the corner $h$ at one of its appearances is shorter than $\alpha^*$, belongs to $\mathbf{C}_M$, and has the same convex hull as $\alpha^*$. Since two arcs with the same convex hull have the same width, this contradicts the minimality of $\alpha^*$.

*Proof of* (ii). Each segment endpoint is in $H$ because if $\alpha^*$ had a segment endpoint not in $H$, a shorter arc with fewer segments and the same convex hull could be made by cutting the corner just as in (i). □

LEMMA 3. *A minimal arc $\alpha^*$ has no self-intersections.*

*Proof.* Lemma 2 implies that if $\alpha^*$ does cross itself, it must do so strictly inside its convex hull. Suppose there is such a crossing. Then there are points $D$, $E$, $F$, $G$ along the arc, visited in that order, such that segment $DE$ of the arc intersects segment $FG$. Thus

$$A\tilde{}B = A\tilde{}DE\tilde{}FG\tilde{}B,$$

where $A\tilde{}D$ or $G\tilde{}B$ might be trivial. Now $DE$ and $FG$ are the diagonals of a quadrilateral, of which $DF$ and $EG$ are two opposite sides. So by adding two applications of the triangle inequality

$$|DF| + |EG| < |DE| + |FG|.$$

Consider the new arc

$$A\tilde{}DF\tilde{}EG\tilde{}B,$$

where $F\tilde{}E$ is $E\tilde{}F$ reversed. The convex hull of the new arc is the same as that of $\alpha^*$, but by the above inequality the new arc is shorter, contradicting the minimality of $\alpha^*$. □

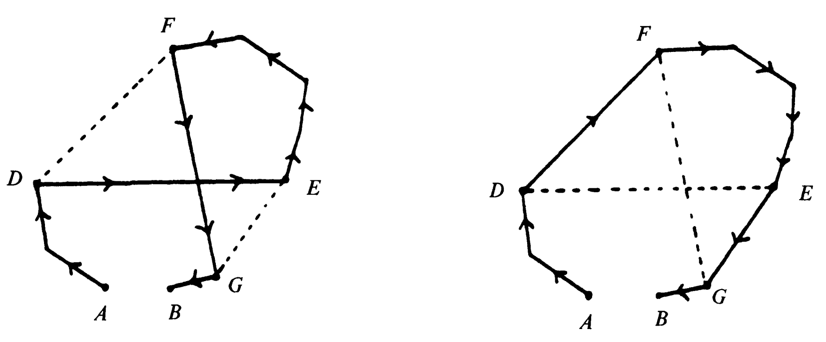



FIG. 14. Construction for the proof of Lemma 3.

According to the above, a minimal arc $\alpha^*$ must visit each corner of its convex hull exactly once without ever intersecting itself, by a succession of line segments between corners of the convex hull. For want of a better term let us call such an arc *standard*. The initial segment of a standard arc must be an edge of its convex hull, to avoid self-intersection. The simplest standard arcs are convex arcs forming all but one side of a convex polygon. We say such an arc has *one gap*. Then there are arcs of standard form with *two or more gaps*:

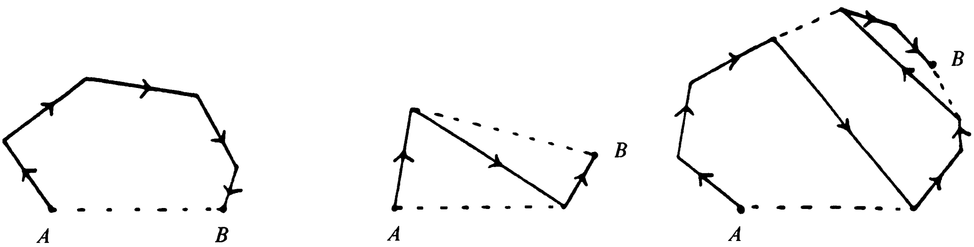



FIG. 15. Standard arcs with 1, 2 and 3 gaps.

In general, an arc of standard form must comprise all segments of its convex hull except for a finite number of these segments, which we call *gaps*.

LEMMA 4. *A minimal arc $\alpha^*$ of width at least 1 has endpoints at distance $d(\alpha^*) \geqslant 1$.*

*Proof.* Let $\alpha^*$ have endpoints $A$ at the origin and $B$ on the positive $x$-axis in $x$, $y$ coordinates. Let $CAD$ be the corner of the convex hull of $\alpha^*$ at $A$, $EBF$ the corner at $B$. Because of the standard form of $\alpha^*$, one of the segments $AC$ and $AD$ must be a segment of $\alpha^*$, while the other must be a gap on the convex hull; similarly for $BE$ and $BF$. We will show that angles $CAD$ and $EBF$ are both at most right angles.

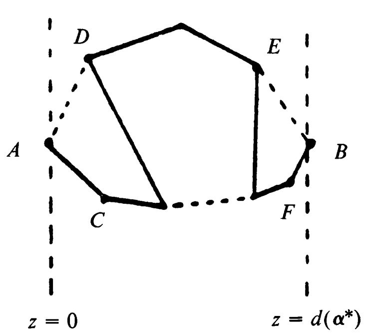



FIG. 16. Construction for proof of Lemma 4.

Once this is proved, it is clear that $\alpha^*$ must lie within the strip bounded by the lines $x = 0$ and $x = d(\alpha^*)$. Since the width of $\alpha^*$ in the horizontal direction is at least 1, $d(\alpha^*)$ must be at least 1.

Suppose angle $CAD$ is obtuse. Assume to be definite that $AC$ is the first segment of $\alpha^*$, so $AD$ is a gap. Let $A'$ be the foot of the perpendicular dropped from $C$ to $\overline{AD}$. Let $C^{\sim}B$ denote the portion of $\alpha^*$ starting at $C$ and ending at $B$. Consider the new arc $\alpha^{**} = A'C^{\sim}B$. This arc $\alpha^{**}$ is clearly shorter than $\alpha^*$. But its convex hull contains that of $\alpha^*$, so its width is at least 1. This contradicts the minimality of $\alpha^*$.

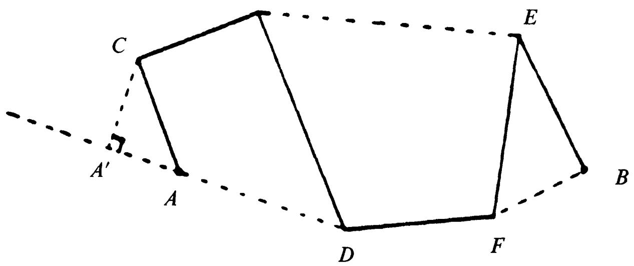



FIG. 17. Construction in case angle $CAD$ is obtuse.

Thus angle $CAD$ is at most a right angle. A similar argument shows angle $DBE$ is at most a right angle, completing the proof of the Lemma. □

Now by Proposition 1, no matter what the bound on the number of segments, for a minimal arc $\alpha^*$ of width at least 1,

$$l(\alpha^*) \geqslant \sqrt{4 + d^2(\alpha^*)} \geqslant \sqrt{5} = 2.236\ldots.$$

Approximating any arc by arcs with a finite number of straight segments, it follows that every arc of width at least 1 has length at least $\sqrt{5}$. But more work is required to get up to the length 2.278... of the caliper of Section 2. This is the subject of the next section.

**4. Convexity of minimal arcs.** It will be shown in this section that for any positive integer $M$, a minimal arc $\alpha^*$ of width at least 1 with at most $M$ segments must be convex. It follows that the caliper defined in Section 2 is in fact a shortest arc among all arcs with width at least 1. According to the previous section $\alpha^*$ is *standard*. That is to say, it must visit each corner of its convex hull exactly once, without ever intersecting itself, by a succession of line segments between corners of the convex hull. In general, a standard arc must pass along each edge of its convex

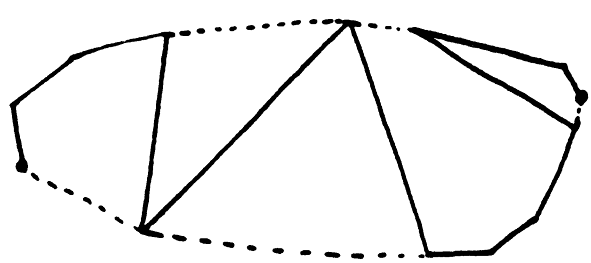


FIG. 18. A standard arc with 5 gaps.

hull except for a finite number of edges which we call gaps. Assuming that the arc starts clockwise from $A$, it must wind clockwise around the hull until it meets a gap. This might be the end, in which case the arc is convex with just one gap. Otherwise, if the arc has $n$ gaps it must cross back and forth across the interior of the hull by say $n-1$ straight segments, winding successively anticlockwise and clockwise around the hull between these crossings.

Associated with each gap of a standard arc is a convex portion of the arc which connects the ends of the gap. Call this portion of arc the *arch* across the gap. A standard arc thus consists of a succession of arches across gaps, with adjacent arches overlapping by a segment crossing the hull. Define the *altitude* of an arch to be the width of the arch between parallels to its gap. A key step in our argument is provided by the following lemma.

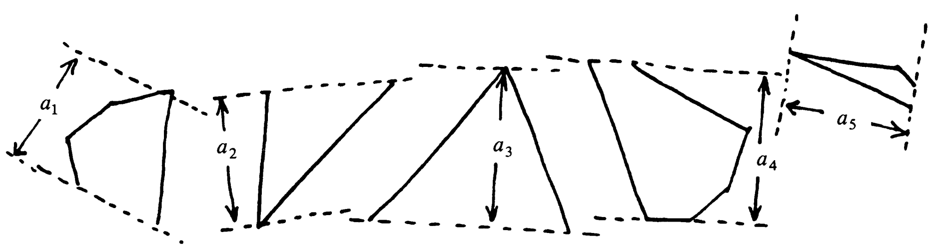



FIG. 19. The arches of the arc in FIG. 18 and their altitudes $a_i$.

LEMMA 5. *If a standard arc has width $w$, at least one of its arches has altitude at least $w$.*

*Proof.* For a convex arc with one gap this is trivial. Consider then an arc with two arches of altitudes $a_1$ and $a_2$:

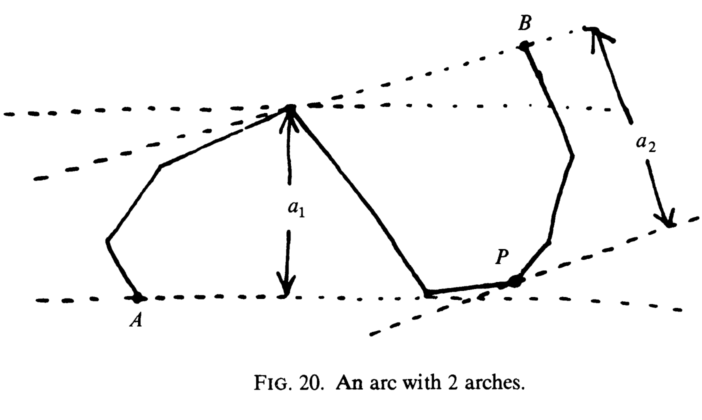



FIG. 20. An arc with 2 arches.

The first arch is taken to be above the first gap, which is assumed horizontal. If $a_1 \geqslant w$ there is nothing to prove. So suppose $a_1 < w$. Then to achieve width $w$ between horizontals the end of the second gap at $B$ must lie above the horizontal at level $a_1$. The second gap must therefore have positive slope. By convexity, the

parallel to this gap supporting the arc from below must touch the arc at a point $P$ on the second arch. Consequently, this arch has altitude $a_2 \geqslant w$.

This argument extends to a standard arc with $n$ arches as follows. Keep track of the order in which the arches are traversed. Suppose that the supporting parallel opposite a gap touches the arc at a point of the arch over that gap, or on some later arch. Then if that point of contact is over the arch, the altitude of the arch must be at least $w$. And in case that point of contact is not on the arch over the gap, but later, the supporting parallel opposite the next gap touches the arc on or after the next gap. This is shown by the same argument as above in case $n = 2$. This serves as an inductive step which forces some arch to have altitude at least $w$, because there are only a finite number $n$ of arches. □

LEMMA 6. *A minimal arc $\alpha^*$ with at most $M$ segments is convex.*

*Proof.* For convenience, rotate and reflect $\alpha^*$ so that the gap corresponding to the tallest arch is horizontal. By Lemma 5, this arch has altitude at least 1. Let $C_1$ and $C_2$ denote the left and right endpoints of the gap.

Suppose $C_2$ is not an endpoint of the arc. Then there is a last point of the arc on the convex hull before $C_2$, call it $L_1$, which is the end of another gap, say $L_1L_2$. It is clear that $\overline{L_1L_2}$ cannot be parallel to $\overline{C_1C_2}$ since if it were the length of $\alpha^*$ would be at least 3. Thus $\overline{L_1L_2}$ must meet $\overline{C_1C_2}$ at some point $E$, say, to the right of $C_2$, as in the figure below. This point $E$ must also be to the right of a perpendicular dropped vertically from $L_1$, since otherwise this perpendicular would offer a shorter path with bigger convex hull.

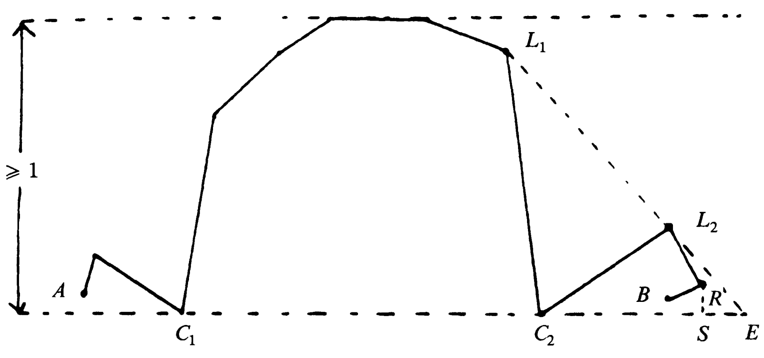



FIG. 21. Case (i).

*Case* (i). $|L_1L_2| < |L_1C_2|$. In this case we could replace the portion $L_1 \tilde{} B$ of $\alpha^*$ by the path $L_1L_2 \tilde{} RS$, as in FIG. 21 where $R$ is the rightmost point of $\alpha^*$, and $S$ is the foot of a perpendicular dropped from $R$ to $C_1C_2$. But since $|RS| \leqslant |C_2L_2|$ this gives a shorter arc with the same convex hull.

*Case* (ii). $|L_1L_2| \geqslant |L_1C_2|$. Now $L_2$ must lie on or outside the circle with center $L_1$ and radius $|L_1C_2|$, as in FIG. 22. Let $L_1L_2$ intersect the circle at $D$, and let the tangent to the circle at $D$ intersect $C_1C_2$ at $F$.

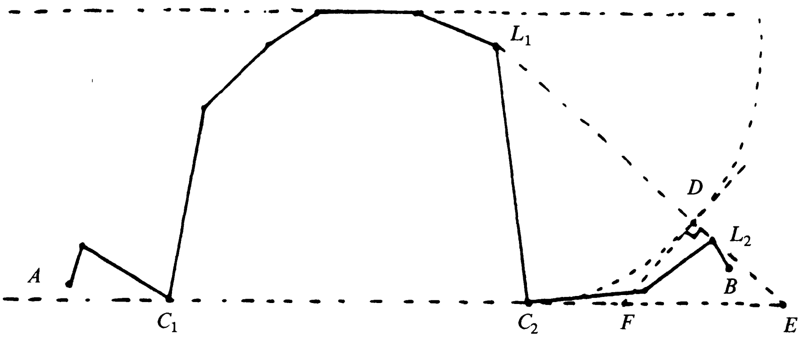



FIG. 22. Case (ii).

By the geometry of the diagram

$$|DF| < |FL_2| < |C_2L_2| < l(C_2\tilde{\ }L_2) \leqslant |DE|.$$

The last inequality is due to the minimality of $\alpha^*$, since otherwise we could replace the portion $L_1\tilde{\ }L_2$ of $\alpha^*$ by the two straight line segments $LD$ and $DE$ to get a shorter arc with a bigger convex hull. Hence $|DF| < |DE|$. That is,

$$\text{angle } C_2EL_1 = \text{angle } FED < 45^\circ.$$

Let $T$ denote a point on the arch from $C_1$ to $C_2$ farthest away from the gap, so the perpendicular distance between $T$ and $\overline{C_1C_2}$ is the altitude of the arch. Reflect $T$ about $C_1C_2$, and do the same with $L_2$. Let $L_2'$ be the reflection of $L_2$, and $T'$ the reflection of $T$.

Then

$$\begin{aligned} l(T\tilde{\ }C_2) + l(C_2\tilde{\ }L_2) &\geqslant |TC_2| + |C_2L_2| \\ &= |TC_2| + |C_2L_2'| \\ &\geqslant \textit{distance of } T \textit{ from } \overline{T'E}. \end{aligned}$$

This distance is least when angle $C_2EL_2$ is $45^\circ$, that is, angle $TET' = 90^\circ$. And in

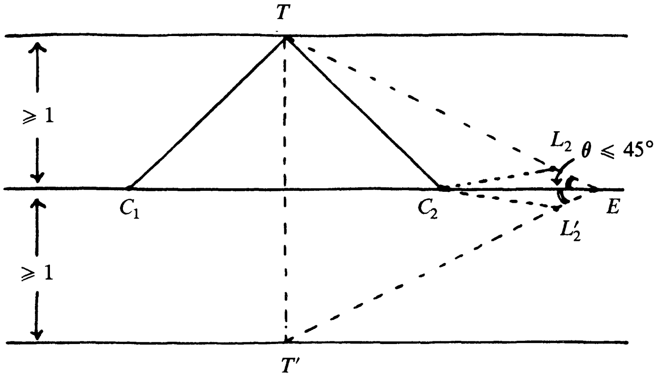



FIG. 23. The reflection argument.

this case the distance is at least $\sqrt{2}$. So

$$l(\alpha^*) \geqslant l(C_1\tilde{\ }T\tilde{\ }C_2\tilde{\ }B) \geqslant 1 + \sqrt{2} > 2.4.$$

But by Lemma 1, $l(\alpha^*)$ must be less than 2.309... . This contradicts the minimality of $\alpha^*$, so $C_2$ must be an end of the arc.

Equally, $C_1$ must be the other end of the arc. That is to say, $\alpha^*$ is a convex arc with a single arch. □

## 5. Exercises and Open Problems.

*Uniqueness of the shortest arc of width one.* A weakness of the indirect argument of the last section is that it does not seem to settle the uniqueness question. While the arguments of section 3 show that the caliper is unique among convex arcs, and some variation of the argument of the last section should ensure uniqueness among arcs that cross their convex hull only a finite number of times, what about arcs that cross an infinite number of times? It is hard to imagine such an arc of width one of the same length as the caliper, but we do not have a proof that no such arc exists.

*Maximizing the probability of crossing.* Consider again the problem, posed in the introduction, of maximizing the probability that a dropped piece of wire of length $l$ will cross parallel lines distance 1 apart. In case the arc is not planar there may be some ambiguity about what constitutes a crossing. But however this ambiguity is resolved, it is obviously best to keep the wire bent in a planar arc. If the arc has width $w_\theta$ between parallels at angle $\theta$ to the $x$-axis, then by conditioning on the angle $\theta$ at which the arc falls relative to the parallels, the chance of crossing is

$$P(cross) = (1/\pi)\int_0^\pi \min(w_\theta, 1)\, d\theta.$$

The problem is to maximize this integral over all arcs of length $l$. Call the maximal value of $P(cross)$ so obtained $P^*(l)$. The existence of an arc of length $l$ attaining this bound can be shown by the usual kind of compactness argument. For $l \leqslant 1$, the optimal arc is straight. And for $l > 2.27829\ldots$, the length of the caliper, $P^*(l)$ is 1. But for $l$ between 1 and the length of the caliper, we do not know how to find an optimal arc, whether it is unique, or what is the value of $P^*(l)$.

*Upper bounds on $P^*(l)$.* Because the expected number of crossings is $(2/\pi)l$, no matter what the arc, Markov's inequality gives

$$P^*(l) \leqslant (2/\pi)l$$

The graph of the right-hand side as a function of $l$ is the straight line through the origin in FIGURE 25. For $l \leqslant 1$ this Markov bound is attained. And for $l$ greater than the length of the caliper, $P^*(l)$ is 1. But we do not know any better upper bounds on $P^*(l)$ than 1 and the Markov bound.

*Lower bounds on $P^*(l)$.*

1. *Leaving the arc straight.* Lower bounds on $P^*(l)$ are of course easier to obtain, by computation of $P(cross)$ for particular arcs of length $l$. For a straight arc of length $l \geqslant 1$ the crossing probability is well known and easily calculated:

$$P_{straight}(l) = (2/\pi)\big(l(1 - \sin\beta) + \beta\big), \quad \text{where } \beta = \arccos(1/l).$$

The graph of $P_{straight}(l)$ is the lowest curve in FIGURE 25.

2. *Bending the arc in the middle.* The next easiest case is an arc of two straight segments bent in the middle. Let $d$ be the distance between the ends, $h = l/2$, the length of each segment, as in FIGURE 24.

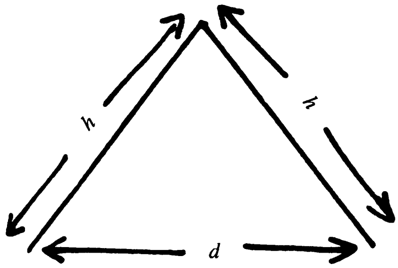



FIG. 24. A straight arc bent in the middle.

In case $d \leqslant 1$ and $h \leqslant 1$, when the triangle defined by the arc must cross the lines either zero or two times, the crossing probability is simply

$$(2/\pi)(\textit{perimeter of triangle})/2 = (l + d)/\pi,$$

since the expected number of crossings is $(2/\pi)(\textit{perimeter of triangle})$. So for fixed $1 \leqslant l \leqslant 2$ it is best to make $d \geqslant 1$. According to our calculations, which we invite the reader to check as an exercise, for $l$ between 0 and $2/\sqrt{3}$ it is best to leave the arc straight. For $l > 2/\sqrt{3}$, the arc should be bent in the middle so that the distance $d$ between its ends equals $2/\sqrt{3}$. Let $P^*_{middle}(l)$ denote the maximum crossing probability for an arc of length $l$ bent in the middle. Then

$$\begin{aligned}
P^*_{middle}(l) &= P_{straight}(l) \quad \text{for } 0 \leqslant l \leqslant 2/\sqrt{3}, \\
&= \frac{l}{\pi} + \frac{1}{3} \quad \text{for } 2/\sqrt{3} < l \leqslant 2, \\
&= \frac{l}{\pi} + \frac{1}{3} - \frac{4}{\pi}\left[\sqrt{h^2 - 1} - \arccos 1/h\right] \quad \text{for } 2 < l < 4/\sqrt{3}, \\
&= 1 \text{ for } l \geqslant 4/\sqrt{3}.
\end{aligned}$$

The function $P^*_{middle}(l)$ is the middle curve in FIGURE 25. By the formula above, $P^*_{middle}(2.27829\ldots) = .99794\ldots$. So $P^*(l)$ is very close to $P^*_{middle}(l)$ for $l$ near the length of the caliper.

3. *Bending the arc into two straight segments.* We conjecture that $P^*_{middle}(l)$ is in fact the maximum crossing probability over all arcs of length $l$ bent, not necessarily in the middle, into two straight segments. We can calculate the crossing probability of any arc bent into two straight pieces. But the formula for the probability is forbidding, and we have not gone through the calculus necessary to verify our conjecture. To see what the formula is like, consider the triangle of FIG. 26.

The crossing probability is the integral

$$\int_0^{\pi} \min(w_\theta, 1)\, d\theta.$$

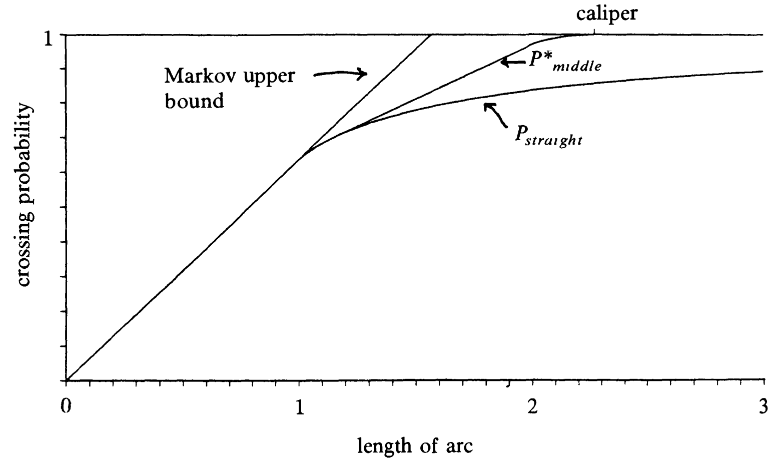



FIG. 25. Crossing probabilities for various arcs.

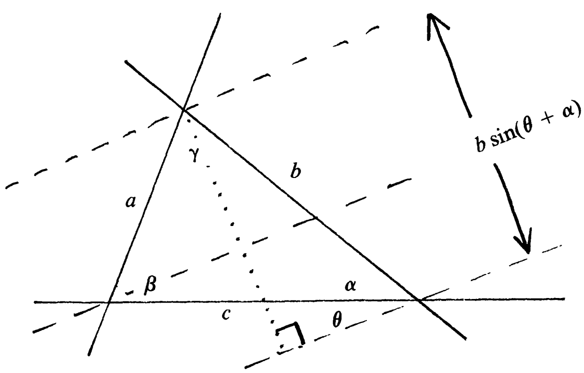



FIG. 26. Triangle defined by into two straight segments.

divided by $\pi$. As can be seen from the diagram, this integral breaks into three parts, each representing the contribution from an interval of angles $\theta$ for which the width is attained by a particular side of the triangle. The length of this interval is the angle opposite the side. Here is the contribution from the interval of angles $\theta$ from 0 to $\beta$ in FIGURE 26 for which side $b$ attains the width, call it $Int(\beta)$. The contributions of angles $\alpha$ and $\gamma$ are similar, with cyclic changes in sides and angles.

$$Int(\beta) = \int_0^\beta \min(b\sin(\theta + \alpha), 1)\, d\theta.$$

$$= b(\cos\gamma + \cos\alpha) \qquad \text{if } b < 1.$$

For $b \geqslant 1$, the integral is more complicated, but reduces after some calculation, as follows. Let

$$\Phi_+ = (\pi/2 - \alpha) + \arccos 1/b,$$

$$\Phi_- = (\pi/2 - \alpha) - \arccos 1/b.$$

For a function $f$ of $\theta$, let $[f(\theta)]_x^y$ denote $f(y) - f(x)$. Then,

$$\begin{aligned}
Int(\beta) &= [-b\cos(\alpha+\theta)]_0^{\Phi_-} + 2\arccos(1/b) + [-b\cos(\alpha+\theta)]_{\Phi_+}^{\beta} \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{if } 0 < \Phi_- < \Phi_+ < \beta \\
&= \Phi_+ + [-b\cos(\alpha+\theta)]_{\Phi_+}^{\beta} && \text{if } \Phi_- < 0 < \Phi_+ < \beta \\
&= [-b\cos(\alpha+\theta)]_0^{\Phi_-} + (\beta - \Phi_-) && \text{if } 0 < \Phi_- < \beta < \Phi_+ \\
&= \beta && \text{if } \Phi_- < 0 < \beta < \Phi_+.
\end{aligned}$$

Add this expression and two similar expressions for $Int(\alpha)$ and $Int(\gamma)$, and divide by $\pi$ to get the crossing probability for an arc composed of two straight segments. The reader is invited to check this analysis, and verify our conjecture that bending in the middle always maximizes this crossing probability for given length.

Let $P_{middle}(l, d)$ denote the chance that an arc of length $l$ bent in the middle into two segments, with endpoints distance $d$ apart, crosses at least once. It is interesting to note that in the case $l \leqslant 2$, $d \geqslant 1$, it is possible to bend the arc at points other than the center and still have chance $P_{middle}(l, d)$ of crossing. Suppose the arc is bent so that the distance between the two endpoints is still $d$, but the two arms have different lengths. Then provided both arms have length less than one, the chance that the arc crosses at least once is $P_{middle}(l, d)$. To see this, consider the triangular closed arc. This crosses either zero, two, or four times. Since the distance between the ends is fixed, the perimeter of the triangle is the same no matter where the arc is bent. So the expected number of crossings is the same, no matter where the arc is bent. And if both arms have length less than 1, then four crossings are made if and only if the base of the triangle crosses twice. So the probability of crossing four times is the same no matter where the arc is bent, provided both arms have length less than 1. This implies the chance of at least two crossings remains the same. Thus if our conjecture above is correct, then in the case $l < 2$ there are infinitely many ways to bend the arc in two to maximize the probability of crossing.

4. *Bending into 3 or more segments.* This will, of course, lead to better lower bounds on $P^*(l)$. But after our calculations above for two segments we do not see how to organize the computations sensibly, let alone evaluate $P^*(l)$ as a limit. All we know for sure is that this maximal probability as a function of $l$ lies somewhere in between the top two curves in FIGURE 25. Once again, the reader is invited to help.

*A variation of the problem.* A nice conjecture about a variation of the problem was suggested to us by David Goldschmidt. Suppose you are allowed to cut the wire and glue the pieces back together to form a connected union of arcs, call it a *gadget*. How now can you improve the crossing probability, and what is the shortest length from which you can make a gadget of width 1? The conjecture for the shortest gadget of width 1 is three straight arms connecting the corners of an equilateral triangle of altitude 1 to its center. Since this gadget is so much simpler to define than the caliper, perhaps there is a simple proof.

*The problem in higher dimensions.* What is the shortest arc of width one in $d$ dimensions? This problem seems very much harder in three or more dimensions. The three-dimensional problem can be presented as follows. Suppose there is a one

inch gap between the parallel plane sides of your stove and your kitchen cupboard. What is the shortest length of a piece of wire that can be bent in such a way that it cannot fall into the crack? Presumably the solution is shaped something like three connected sides of a regular tetrahedron of altitude one inch, but we have no idea of what the exact shape must be.

REFERENCES

1. M. E. Barbier, Note sur le problème de l'aiguille et jeu du joint couvert, *J. Mathém. Pures et Appl.*, (2) 5 (1860) 272–286.
2. Duane W. DeTemple and Jack M. Robertson, Constructing Buffon curves from their distributions, *Amer. Math. Monthly*, 87 (1980) 779–784.
3. N. D. Kazarinoff, Geometric Inequalities, Random House, 1961.
4. I. Niven, Maxima and Minima Without Calculus, Dolciani Math. Expos. #6, Math. Assoc. Amer., 1982.
5. I. M. Yaglom and V. G. Boltyanskii, Convex Figures, Holt, Rinehart and Wilson, 1961.

# The Use of Full Covers in Real Analysis

MICHAEL W. BOTSKO, *St Vincent College, Latrobe, PA*

MICHAEL W. BOTSKO: I received my B.A. and M.A. from Duquesne University and my Ph.D. from the University of Pittsburgh under the supervision of George Laush. My main interest is the teaching of mathematics, but I also enjoy studying in the area of real functions.

**1. Introduction.** It is the purpose of this article to show how the concept of a full cover can be used to simplify and unify the proofs of some harder theorems in elementary analysis. Let us begin with the definition of a full cover which was given in [**6**].

DEFINITION. *Let $[a, b]$ be a given closed, bounded interval and let $X$ be a subset of $[a, b]$. A collection $\mathscr{C}$ of closed subintervals of $[a, b]$ is a full cover of $X$ if to each $x$ in $X$ there corresponds a number $\delta(x) > 0$ such that every closed subinterval of $[a, b]$ that contains $x$ and has length less than $\delta(x)$ belongs to $\mathscr{C}$.*

Since a collection $\mathscr{C}$ of closed intervals is a *Vitali cover* of a set $X$ if at each point $x$ in $X$ there are in $\mathscr{C}$ intervals containing $x$ of arbitrarily small length, it is clear that a full cover of $X$ is also a Vitali cover but not conversely. For Vitali covers we have the well known and rather hard to prove Vitali Covering Theorem. For full covers, however, the following much easier result is available to us; see [**6**, p. 79].

THOMSON'S LEMMA.* *If $\mathscr{C}$ is a full cover of $[a, b]$, then $\mathscr{C}$ contains a partition of $[a, b]$, i.e., there is a partition of $[a, b]$ all of whose closed subintervals belong to $\mathscr{C}$.*

*Proof.* (The proof is given in [**1**] and [**6**] and is repeated here for convenience.) Suppose $\mathscr{C}$ contains no partition of $[a, b]$. Then by repeated bisection of $[a, b]$, there must be a sequence $\{J_n\}$ of closed subintervals of $[a, b]$ such that $J_n \supseteq J_{n+1}$ for each $n$, $|J_n| \to 0$, and $\mathscr{C}$ contains no partition of any $J_n$. ($|J_n|$ denotes the length of $J_n$.) By the Nested Interval Theorem there exists an $x$ in the intersection of the sequence $\{J_n\}$. Consider $\delta(x)$ as is given by the above definition. Since $|J_n| \to 0$ there exists a positive integer $N$ so large that $|J_N| < \delta(x)$. Therefore, $J_N$ belongs to $\mathscr{C}$ whence $\mathscr{C}$ trivially contains a partition of $J_N$, a contradiction.

The idea of extracting partitions from full covers goes back at least as far as 1900, since this technique was used by Goursat in [**3**] to prove the Cauchy Integral Theorem. More recently (see [**6**]), Brian S. Thomson used the method to prove certain monotonicity theorems involving ordinary derivatives. For an application of Thomson's Lemma to some easier proofs in elementary analysis, see [**1**].

*It has recently come to my attention that this type of result was published by Pierre Cousin in *Acta Math.*, vol. 19 (1895) 22.

Let us close this section with two easy remarks that will facilitate some of the proofs in this article.

*Remark* 1. Let $E$ and $F$ be subsets of $[a, b]$. If $\mathscr{E}$ is a full cover of $E$ and $\mathscr{F}$ is a full cover of $F$, then $\mathscr{E} \cup \mathscr{F}$ is a full cover of $E \cup F$.

The proof is immediate.

*Remark* 2. Let $X$ be a subset of $[a, b]$ and let $\mathscr{G}$ be a collection of open sets that covers $X$. If $\mathscr{C}$ is the collection of closed subintervals of $[a, b]$ each of which is contained in some set in $\mathscr{G}$, then $\mathscr{C}$ is a full cover of $X$.

*Proof.* Let $x$ belong to $X$ which implies that $x$ is in $O$ for some $O$ in $\mathscr{G}$. Thus there exists $\delta(x) > 0$ such that $(x - \delta(x), x + \delta(x)) \subseteq O$. Now let $I$ be any closed subinterval of $[a, b]$ that contains $x$ and has length less than $\delta(x)$. Since $I \subseteq (x - \delta(x), x + \delta(x)) \subseteq O$, it follows that $I$ belongs to $\mathscr{C}$ and that $\mathscr{C}$ is a full cover of $X$.

**2. A related method.** Recently Professor Leonard Gillman informed the author that an idea related to the lemma was used by L. R. Ford in **[2]** and again by P. Shanahan in **[5]** to prove some theorems in elementary analysis. In order to discuss the method given by Ford and Shanahan we need two definitions.

Let $\mathscr{C}$ be a collection of subsets of $[a, b]$. We say that $\mathscr{C}$ is *local* if each $x \in [a, b]$ has a neighborhood, with respect to the relative topology on $[a, b]$, which is a member of $\mathscr{C}$. We say that $\mathscr{C}$ is *additive* if whenever $C_1$ and $C_2$ are members of $\mathscr{C}$ such that $C_1 \cap C_2 \neq \varnothing$, then $C_1 \cup C_2 \in \mathscr{C}$.

We now state the main theorem used by Shanahan in **[5]** and give a proof of this theorem using Thomson's Lemma.

THEOREM 1. *If $\mathscr{C}$ is a local, additive collection of closed subintervals of $[a, b]$, then $\mathscr{C}$ contains $[a, b]$.*

*Proof.* Let
$$\mathscr{D} = \{I\colon I \subseteq J \text{ for some } J \text{ in } \mathscr{C}\}.$$
(In this and all other proofs, $I$ represents a closed subinterval of $[a, b]$.) Clearly $\mathscr{D}$ is a full cover of $[a, b]$ and by Thomson's Lemma there exists $P = \{x_0, x_1, x_2, \dots, x_n\}$ a partition of $[a, b]$ such that each $I_k = [x_{k-1}, x_k]$ belongs to $\mathscr{D}$. Now to each $I_k$ there corresponds a $J_k$ in $\mathscr{C}$ such that $I_k \subseteq J_k$. Since $[a, b]$ is the union of $J_1, J_2, \dots, J_n$ and $\mathscr{C}$ is additive, we clearly have that $[a, b]$ belongs to $\mathscr{C}$.

In a letter to the author, Professor Gillman restated the above theorem in the following way.

THEOREM 1 (2nd version). *If $\mathscr{C}$ is a local collection of closed subintervals of $[a, b]$, then $\mathscr{C}$ contains a finite subcover of $[a, b]$.*

Because of this Professor Gillman made the observation that Theorem 1 actually paraphrases the Heine-Borel Theorem.

When we consider the second version of Theorem 1, we see the advantage in using Thomson's Lemma for proving certain theorems in analysis. The advantage of course is that Thomson's Lemma gives a finite subcover consisting of abutting closed intervals. Because of this the lemma is more appropriate than Theorem 1 for the proofs that we will be considering in this article.

**3. Proving theorems using full covers.** Let us now begin our applications of the lemma by presenting one of the monotonicity theorems given by Thomson in **[6]**.

THEOREM 2. *If $\underline{D}f(x) \geqslant 0$ everywhere on an interval $[A, B]$, then $f$ is nondecreasing on that interval. ($\underline{D}f(x)$ is the lower derivative of $f$ at $x$.)*

*Proof.* Let $a$ and $b$ belong to $[A, B]$ with $a < b$ and show that $f(a) \leqslant f(b)$. Let $\varepsilon > 0$ be given and let

$$\mathscr{C} = \{I \colon f(I) > -\varepsilon|I|\}.$$

(In the spirit of **[6]**, $f(I)$ denotes $f(d) - f(c)$ where $I = [c, d]$.) To show that $\mathscr{C}$ is a full cover of $[a, b]$ let $x$ belong to $[a, b]$ and note that $\underline{D}f(x) > -\varepsilon$. Thus there exists $\delta(x) > 0$ for which $(f(y) - f(x))/(y - x) > -\varepsilon$ for all $y$ in $[A, B]$ such that $0 < |y - x| < \delta(x)$. Now let $I$ be any closed subinterval of $[a, b]$ such that $x \in I$ and $|I| < \delta(x)$. We can easily show that $f(I) > -\varepsilon|I|$ and hence $I$ belongs to $\mathscr{C}$. Thus $\mathscr{C}$ is a full cover of $[a, b]$ and by Thomson's Lemma there exists a partition $P = \{x_0, x_1, x_2, \ldots, x_n\}$ of $[a, b]$ such that each $I_k = [x_{k-1}, x_k]$ belongs to $\mathscr{C}$. Therefore

$$f(b) - f(a) = \sum_{k=1}^{n} f(I_k) > -\varepsilon \sum_{k=1}^{n} |I_k| = -\varepsilon(b - a).$$

Since $\varepsilon$ is arbitrary we have that $f(b) - f(a) \geqslant 0$ which completes the proof.

By adding additional intervals to his full cover, Thomson also proves more general monotonicity theorems in regard to the ordinary derivative.

Since Thomson's lemma yields a partition of $[a, b]$ whose closed subintervals belong to our full cover, it is ideal for proving Riemann integrability. To illustrate the method we first show that a continuous function is Riemann integrable and then, by slightly modifying our proof, establish a much more general result.

THEOREM 3. *If $f$ is a continuous real valued function on $[a, b]$, then $f$ is Riemann integrable on $[a, b]$.*

*Proof.* Let $\varepsilon > 0$ be given and let

$$\mathscr{C} = \{I \colon f \text{ is bounded on } I \text{ and } \sup_I f - \inf_I f < \varepsilon/(b - a)\}.$$

To show that $\mathscr{C}$ is a full cover of $[a, b]$, let $x$ be any point in $[a, b]$. Since $f$ is continuous at $x$, for $\varepsilon/(4(b - a)) > 0$ there exists $\delta(x) > 0$ for which $|f(y) - f(x)| < \varepsilon/(4(b - a))$ for all $y$ in $[a, b]$ such that $|y - x| < \delta(x)$. Now let $I$ be any closed subinterval of $[a, b]$ with $x \in I$ and $|I| < \delta(x)$. Clearly $f$ is bounded on $I$ and

$$\sup_I f - \inf_I f \leqslant \varepsilon/(2(b - a)) < \varepsilon/(b - a),$$

which says that $I$ is in $\mathscr{C}$. Thus $\mathscr{C}$ is a full cover of $[a, b]$ and by Thomson's Lemma we have $P = \{x_0, x_1, x_2, \ldots, x_n\}$ a partition of $[a, b]$ such that $I_k = [x_{k-1}, x_k]$ is in $\mathscr{C}$ for each $k$. Since $f$ is bounded on each $I_k$, it clearly follows that $f$ is bounded on $[a, b]$. Let us now form the difference between the upper and lower sums of $f$ over the partition $P$.

$$\bar{S}_P(f) - \underline{S}_P(f) = \sum_{k=1}^{n} (M_k - m_k)|I_k|.$$

($M_k = \sup\{f(x)\colon x \in I_k\}$ and $m_k = \inf\{f(x)\colon x \in I_k\}$.)

$$< (\varepsilon/(b-a)) \sum_{k=1}^{n} |I_k| = \varepsilon.$$

Thus $\bar{S}_P(f) - \underline{S}_P(f) < \varepsilon$ and the proof is complete.

We now prove the more general theorem using only a slight modification of the proof of the previous result.

THEOREM 4. *If $f$ is bounded on $[a, b]$ and continuous almost everywhere, then $f$ is Riemann integrable on $[a, b]$.*

*Proof.* There exists $M > 0$ such that $|f(x)| \leqslant M$ for all $x$ in $[a, b]$. Let $E = \{x\colon f \text{ is continuous at } x\}$ and let $\varepsilon > 0$ be given. Since $[a, b] \setminus E$ is of measure zero, for $\varepsilon/4M > 0$ there exists $\{J_n\}$ a sequence of open intervals such that $[a, b] \setminus E \subseteq \cup J_n$ and $\Sigma |J_n| < \varepsilon/4M$. Now let

$$\mathscr{D} = \{I\colon \sup_I f - \inf_I f < \varepsilon/(2(b-a))\}.$$

In addition let

$$\mathscr{E} = \{I\colon I \subseteq J_n \text{ for some } n\}.$$

Reasoning as in the previous proof, we have that $\mathscr{D}$ is a full cover of $E$. By Remark 2, $\mathscr{E}$ is a full cover of $[a, b] \setminus E$. Thus by Remark 1, $\mathscr{C} = \mathscr{D} \cup \mathscr{E}$ is a full cover of $[a,b]$ and there is a partition $P = \{x_0, x_1, x_2, \ldots, x_n\}$ of $[a, b]$ such that each $I_k = [x_{k-1}, x_k]$ is in $\mathscr{C}$. Once again we consider the difference between the upper and lower sums of $f$ over the partition $P$.

$$\bar{S}_P(f) - \underline{S}_P(f) = \sum_{(\mathscr{D})} (M_k - m_k)|I_k| + \sum_{(\mathscr{E})} (M_k - m_k)|I_k|.$$

(The first summation is taken over subintervals from $\mathscr{D}$ and the second over subintervals from $\mathscr{E}$.)

$$< (\varepsilon/(2(b-a))) \sum_{(\mathscr{D})} |I_k| + 2M \sum_{(\mathscr{E})} |I_k|$$

$$\leqslant (\varepsilon/(2(b-a)))(b-a) + 2M \sum |J_n|$$

$$< \varepsilon/2 + 2M(\varepsilon/4M) = \varepsilon.$$

Therefore, $f$ is Riemann integrable on $[a, b]$.

Thus Theorem 4 which is usually thought of as much more difficult than Theorem 3, is proved with very little additional work by using the method of full covers.

We now want to give an elementary proof that an absolutely continuous function whose derivative vanishes a.e. is constant. Once again we illustrate the idea by proving a much easier result and then, by slightly modifying our proof, establish the theorem.

THEOREM 5. *If $f'(x) = 0$ for all $x$ on $[a, b]$, then $f$ is constant on $[a, b]$.*

*Proof.* Let $\varepsilon > 0$ be given and let $\mathscr{C} = \{I\colon |f(I)| < \varepsilon |I|\}$. (Recall that $f(I)$ denotes $f(d) - f(c)$ where $I = [c, d]$.) To show that $\mathscr{C}$ is a full cover of $[a, b]$ let $x$ be any point in $[a, b]$. Since $f'(x) = 0$ there exists $\delta(x) > 0$ such that $|(f(y) - f(x))/(y - x)| < \varepsilon$ for all $y$ in $[a, b]$ for which $0 < |y - x| < \delta(x)$. Now if $I$ is

any closed subinterval of $[a, b]$ such that $x \in I$ and $|I| < \delta(x)$, it is easy to see that $|f(I)| < \varepsilon|I|$ so that $I$ belongs to $\mathscr{C}$. Therefore, $\mathscr{C}$ is a full cover of $[a, b]$ and by Thomson's Lemma there exists a partition $P = \{x_0, x_1, x_2, \dots, x_n\}$ of $[a, b]$ such that each $I_k = [x_{k-1}, x_k]$ belongs to $\mathscr{C}$. Therefore,

$$|f(b) - f(a)| \leqslant \sum_{k=1}^{n} |f(I_k)| < \varepsilon \sum_{k=1}^{n} |I_k| = \varepsilon(b-a).$$

Since $\varepsilon$ is arbitrary we have that $f(b) = f(a)$ and it follows that $f(x) = f(a)$ for all $x$ in $[a, b]$.

Let us now prove the much more general result which is often established by using the Vitali Covering Theorem; see [**4**, pp. 90–91].

THEOREM 6. *If $f$ is absolutely continuous on $[a, b]$ and if $f'(x) = 0$ almost everywhere, then $f$ is constant on $[a, b]$.*

*Proof.* Let $E = \{x\colon x \in [a, b] \text{ and } f'(x) = 0\}$ and let $\varepsilon > 0$ be given. Since $f$ is absolutely continuous on $[a, b]$, there exists $\delta > 0$ such that $\sum_{k=1}^{n}|f(b_k) - f(a_k)| < \epsilon$ for every finite collection $\{[a_k, b_k]\}$ of nonoverlapping subintervals of $[a, b]$ with $\sum_{k=1}^{n}(b_k - a_k) < \delta$. Since $[a, b] \setminus E$ is of measure zero there exists $\{J_n\}$ a sequence of open intervals such that $[a, b] \setminus E \subseteq \cup J_n$ and $\sum|J_n| < \delta$. Now let

$$\mathscr{D} = \{I\colon |f(I)| < \varepsilon|I|\} \text{ and } \mathscr{E} = \{I\colon I \subseteq J_n \text{ for some } n\}.$$

$\mathscr{D}$ is a full cover of $E$ by the previous proof and $\mathscr{E}$ is a full cover of $[a, b] \setminus E$ by Remark 2. Hence $\mathscr{C} = \mathscr{D} \cup \mathscr{E}$ is a full cover of $[a, b]$ by Remark 1. Thus by the lemma there is a partition $P = \{x_0, x_1, x_2, \dots, x_n\}$ of $[a, b]$ such that each $I_k = [x_{k-1}, x_k]$ belongs to $\mathscr{C}$. Clearly

$$|f(b) - f(a)| \leqslant \sum_{(\mathscr{D})} |f(I_k)| + \sum_{(\mathscr{E})} |f(I_k)|.$$

(The first summation is taken over subintervals from $\mathscr{D}$ and the second over subintervals from $\mathscr{E}$.)

$$< \varepsilon \sum_{(\mathscr{D})} |I_k| + \varepsilon$$

(This is true because $\sum_{(\mathscr{E})}|I_k| \leqslant \sum|J_n| < \delta$.)

$$\leqslant \varepsilon(b - a + 1).$$

Therefore, $|f(b) - f(a)| < \varepsilon(b - a + 1)$ and since $\varepsilon$ is arbitrary, $f(b) = f(a)$. From this it clearly follows that $f(x) = f(a)$ for all $x$ in $[a, b]$ and the proof is complete.

Next we give a proof of the following mean value theorem which requires a somewhat different sort of full cover.

THEOREM 7. *If $f$ is continuous and $g$ is nondecreasing on $[a, b]$ and if $|f'(x)| \leqslant g'(x)$ except for a countable subset of $[a, b]$, then $|f(b) - f(a)| \leqslant g(b) - g(a)$.*

*Proof.* Let $E = \{x\colon x \in [a, b] \text{ and } |f'(x)| \leqslant g'(x)\}$. Then $[a, b] \setminus E$ is countable and can be expressed as $\{y_1, y_2, y_3, \dots\}$. Let $\varepsilon > 0$ be given and let

$$\mathscr{D} = \{I\colon |f(I)| < g(I) + \varepsilon|I|\}.$$

In addition for each positive integer $n$ let

$$\mathscr{E}_n = \{I: I \text{ contains } y_n \text{ and } |f(I)| < \varepsilon/2^{n+1}\}.$$

It is not very difficult to show that $\mathscr{C} = \mathscr{D} \cup (\cup \mathscr{E}_n)$ is a full cover of $[a, b]$. Thus there exists $P = \{x_0, x_1, x_2, \dots, x_m\}$ a partition of $[a, b]$ such that each $I_k = [x_{k-1}, x_k]$ belongs to $\mathscr{C}$. Now

$$|f(b) - f(a)| \leqslant \sum_{(\mathscr{D})} |f(I_k)| + \sum_n \sum_{(\mathscr{E}_n)} |f(I_k)|.$$

(The notation for the sums is similar to that used in the proofs of Theorems 4 and 6.)

$$< \sum_{(\mathscr{D})} (g(I_k) + \varepsilon|I_k|) + \sum_n 2(\varepsilon/2^{n+1})$$

(This is true because there can be at most two closed subintervals of the partition in any $\mathscr{E}_n$.)

$$\leqslant g(b) - g(a) + \varepsilon(b - a) + \varepsilon = g(b) - g(a) + \varepsilon(b - a + 1).$$

Thus, $|f(b) - f(a)| \leqslant g(b) - g(a)$ and the proof is complete.

There are other interesting results along the lines of Theorem 7 that can be proved by Thomson's Lemma. In closing let us state one such theorem.

THEOREM 8. *If $f$ is absolutely continuous on $[a, b]$ and $g$ is nondecreasing and if $|f'(x)| \leqslant g'(x)$ a.e. on $[a, b]$, then $|f(b) - f(a)| \leqslant g(b) - g(a)$.*

(Since the proof is quite easy, we will only give the appropriate full cover and leave the rest to the reader.) Let $\varepsilon > 0$ be given and let $E = \{x: |f'(x)| \leqslant g'(x)\}$. Let $\delta > 0$ be the positive number corresponding to $\varepsilon$ in the definition of absolute continuity of $f$; see the proof of Theorem 6. Since $[a, b] \setminus E$ has measure zero there exists $\{J_n\}$ a sequence of open intervals such that $[a, b] \setminus E \subseteq \cup J_n$ and $\Sigma |J_n| < \delta$. Let

$$\mathscr{D} = \{I: |f(I)| < g(I) + \varepsilon|I|\}$$

and let

$$\mathscr{E} = \{I: I \subseteq J_n \text{ for some } n\}.$$

Then $\mathscr{C} = \mathscr{D} \cup \mathscr{E}$ is a full cover of $[a, b]$ which can easily be used to prove the theorem.

In Theorems 5, 6, 7, and 8 we have assumed that $f$ is a real valued function. It is clear, however, that all four theorems remain valid with essentially the same proofs for $f$ a function from $[a, b]$ into a Banach space.

REFERENCES

1. M. W. Botsko, A unified treatment of various theorems in elementary analysis, this MONTHLY, 94 (1987) 450–452.
2. L. R. Ford, Interval-additive propositions, this MONTHLY, 64 (1957) 106–108.
3. E. Goursat, Sur la definition générale des fonctions analytique, d'après Cauchy, *Trans. Amer. Math. Soc.*, 1 (1900) 14–16.
4. H. L. Royden, Real Analysis, Macmillan, New York, 1963.
5. P. Shanahan, A unified proof of several basic theorems of real analysis, this MONTHLY, 79 (1972) 895–898.
6. B. S. Thomson, On full covering properties, *Real Analysis Exchange*, 6 (1980–81) 77–93.

# UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

*In this department the* MONTHY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

## What Are the Laws of Greed?

JIM PROPP
*University of California, Berkeley, CA 94720*

Given a fixed prime $p$, suppose we partition the natural numbers $0, 1, 2, \ldots$ into disjoint sets $S_0, S_1, S_2, \ldots$, none of which may contain an arithmetic progression of length $p$, and suppose moreover that we effect this partition following the "greedy algorithm." That is, we successively assign each natural number $n$ to the set $S_k$ with $k$ as small as possible, subject to the constraint that this assignment must create no arithmetic progression of length $p$ in $S_k$. What formula governs the partition determined by this algorithm?

To see what sort of formula one might hope for, consider the case $p = 3$ (the only case solved to date aside from the trivial case $p = 2$). Carrying out the greedy algorithm, we find that

$$\begin{array}{ll}
0 \in S_0 & \\
1 \in S_0 & \\
2 \in S_1 & \text{(since } 0,1 \in S_0) \\
3 \in S_0 & \\
4 \in S_0 & \\
5 \in S_1 & \text{(since } 1,3 \text{ (and } 3,4) \in S_0) \\
6 \in S_1 & \text{(since } 0,3 \in S_0) \\
7 \in S_2 & \text{(since } 1,4 \in S_0 \text{ and } 5,6 \in S_1) \\
8 \in S_2 & \text{(since } 0,4 \in S_0 \text{ and } 2,5 \in S_1)
\end{array}$$

and so on. Gerver, Propp, and Simpson **[1]** define the sequence $(a_n)_0^\infty = (0,0,1,0,0,1,1,2,2,\ldots)$ by the rule that $a_n = k$ exactly if $n \in S_k$, and they prove the recurrence relation

$$a_{3k+r} = \left[\frac{1}{2}(3a_k + r)\right] \quad \text{for } k \geq 0, 0 \leq r \leq 2, \tag{1}$$

where "$[\cdot]$" is the greatest integer function.

The sequence $(a_n)$ can be viewed as an irregular "waltz"

$$(0,0,1;0,0,1;1,2,2;\ldots)$$

played on a semi-infinite piano keyboard. The musical merits of this endless waltz are doubtful, but it *is* easy to memorize: only certain three-note patterns can occur within a measure, and the performer can memorize the sequence of these patterns

for $0 \leq k,\ r \leq p - 1$, where $p$ is any prime. Going on past $n = p^2$, we see that things aren't quite as simple as the $p = 3$ case might lead us to think. It is *almost*, but not quite, true that equation (2) holds for $0 \leq k \leq p^2 - 1$, $0 \leq r \leq p - 1$; extensive computer data suggest that the only exceptions occur at values of $n = pk + r$ of the form $(p^3 - p^2 + p - 1) + m(p - 1)$, with $2 \leq m \leq p - 1$. And even the exceptions have a lawfulness to them: if we define $c_m = a_{(p^3-p^2+p-1)+m(p-1)}$, then we find that $c_m(p)$ depends only on $p$ modulo $m$, thus linking together the behavior of the sequences $a_n(p)$ for all the different possible (prime) values of $p$. Moreover, we find that $c_2, c_3, \ldots, c_{p-1}$ consists of 1's and 2's in precise alternation if and only if $p$ is one of the odd Fermat primes $3, 5, 17, 257, \ldots$.

There is a formula involving the prime factorization of $n - 1$ that appears to accurately describe the behavior of $a_n(p)$ for values of $n$ up to $p^3 - 1$, but certainly one wants much more. Since $S_0(p)$ is understood for all primes $p$ and $S_k(2)$ and $S_k(3)$ are understood for all $k \geq 0$, it is galling that even as "simple" a case as $S_1(5)$ has not yet yielded up the secret of its governing rule. Perhaps readers with a bent for computer calculation and large-scale data analysis will be able to settle this single concrete case; that might prove the opening wedge for the analysis of the general case $S_k(p)$.

REFERENCES

1. Joseph Gerver, James Propp, and Jamie Simpson, Greedily partitioning the natural numbers into sets free of arithmetic progressions, *Proc. Amer. Math. Soc.*, 102 (1988) 765–772.
2. J. L. Gerver and L. T. Ramsey, Sets of integers with no long arithmetic progressions generated by the greedy algorithm, *Math. Comp.*, 33 (1979) 1353–1359.
3. A. M. Odlyzko and R. P. Stanley, Some curious sequences constructed with the greedy algorithm, unpublished Bell Laboratories report, January 1978.
4. B. L. van der Waerden, Beweis einer Baudetschen Vermutung, *Nieuw. Arch. Wisk.*, 15 (1927) 212–216.

# NOTES

EDITED BY DENNIS DETURCK, DAVID J. HALLENBECK, AND ANITA E. SOLOW

## Fair Dice *

PERSI DIACONIS
*Department of Mathematics, Harvard University, Cambridge, MA 02138*

JOSEPH B. KELLER
*Departments of Mathematics and Mechanical Engineering, Stanford University, Stanford, CA 94305*

**1. Introduction.** Dice are usually cubes of a homogeneous material. Symmetry suggests that a homogeneous cube has the same chance of landing on each of its six faces after a vigorous roll, so it is said to be fair. Similarly the four other regular solids—the tetrahedron, octahedron, dodecahedron and icosahedron—are fair. Are there any other fair polyhedra?

To answer this question we must first define what we mean by fair. We shall say that a convex polyhedron is fair by symmetry if and only if it is symmetric with respect to all its faces. This means that any face can be transformed into any other face by a rotation, a reflection, or a combined rotation and reflection, which takes the polyhedron into itself. The collection of all these transformations of a given polyhedron is called its symmetry group. The fact that some transformation in the group takes any given face into any other given face is expressed by saying that the group acts transitively on the faces. Thus we can say that a convex polyhedron is fair by symmetry if and only if its symmetry group acts transitively on its faces.

In the next section we shall determine all such polyhedra. Then in the final section we shall show that there are other polyhedra which are fair, but not fair by symmetry.

**2. Polar reciprocals.** All the symmetry transformations of a convex polyhedron leave invariant its center of gravity, which is a point inside the polyhedron. We form the dual or polar reciprocal polyhedron of the original polyhedron with respect to its center. This is done by passing a plane through each vertex of the original polyhedron, perpendicular to the line from the center to the vertex. These planes form the faces of the polar reciprocal polyhedron, and each of its vertices is called a pole of one face of the original polyhedron.

Because the symmetry group acts transitively on the faces of the original polyhedron, it also acts transitively on their poles, which are the vertices of the polar reciprocal polyhedron. Therefore the polar reciprocal polyhedron is symmetric with respect to its vertices.

The symmetry groups of all polyhedra symmetric with respect to their vertices have been determined [1]. Furthermore for each such group there is one polyhedron

* This research was supported by the Office of Naval Research, the Air Force Office of Scientific Research, the Army Research Office, and the National Science Foundation.

which has regular polygons for faces. These particular polyhedra are the well-known semiregular solids. They comprise thirteen individuals, the Archimedean solids, and the two infinite classes of prisms and anti-prisms, which were recognized as semiregular by Kepler. The regular or Platonic solids are also semiregular.

Now we use the fact that the polar reciprocal of the polar reciprocal of a polyhedron is similar to the original polyhedron. Therefore we have obtained the following result: The polyhedra which are fair by symmetry are duals of the polyhedra symmetric with respect to their vertices. Each symmetry group of a fair polyhedron is represented by a regular solid or the dual of a semiregular solid. Thus in addition to the five regular solids there are thirteen individual polyhedra and two infinite classes among the fair polyhedra.

Lists of the regular and semiregular solids are given on pages 272 and 277 of reference 1, together with some of their properties. From the list on page 272 we see that every semiregular solid has an even number of vertices, so its polar reciprocal has an even number of faces. Therefore every polyhedron fair by symmetry has an even number of faces. Drawings of the semiregular solids are shown on page 280, and they are determined on pages 269–277 and again on pages 279–286. Drawings of the polar reciprocals of the semiregular solids are shown on pages 34 and 35 of Pearce **[2]**, and photographs of models of them are shown on page 54 of Holden **[3]**.

Each of the semiregular solids belongs to a class of polyhedra with the same symmetry group, all of which are symmetric with respect to their vertices. The dual of any one of them is fair by symmetry, but in general it will be less symmetric than the dual of the semiregular solid. For example, suppose that each equilateral triangular face of a semiregular antiprism is replaced by a given isoceles triangle which is not equilateral. The resulting solid is symmetric with respect to its vertices and has the same symmetry group as the semiregular antiprism. Its dual is a fair die with quadrilateral faces which are not symmetric about either of their diagonals. On the other hand, suppose that the square faces of a semiregular prism are replaced by nonsquare rectangles. The new polyhedron is symmetric with respect to its vertices and has the same symmetry group as the semiregular prism. Its dual is again a dipyramid which is fair by symmetry. It differs from the dual of the semiregular prism only in the ratio of the lengths of the sides of a face.

Grünbaum and Shephard **[4]** give a complete classification of the convex polyhedra with symmetries acting transitively on their faces, which they call "isohedra." They are classified by combinatorial isomorphism type and by symmetry group. There are some isohedra for which the positive rotations alone (i.e. those which preserve orientation, so they do not include reflections) do not act transitively on the faces. Earlier Grünbaum **[5]** showed that every isohedron has an even number of faces.

**3. Other fair polyhedra.** There are other fair polyhedra which are not symmetric. To show this we consider, for example, the dual of the $n$-prism, which is a dipyramid with $2n$ identical triangular faces. We cut off its two tips with two planes parallel to the base and equidistant from it. When the cuts are near the tips, the solid has a very small probability of landing on either of the two tiny new faces. However when the cuts are near the base, it has a very high probability of landing on one of them. Therefore by continuity there must be cuts for which the two new faces and the $2n$ old faces have equal probabilities. The locations of those cuts,

which depend upon the mechanical properties of the die and the table, could be found by experiment or by a difficult mechanical analysis along the lines of our previous study of coin tossing [6]. Similar constructions can be carried out starting with other dice which are fair by symmetry, or which are obtained by this construction. We say these dice are fair by continuity.

As an example, let us consider an infinite prism with a regular $n$-gon as its cross section. Let us cut it with two planes a distance $L$ apart perpendicular to its generators, to produce a polyhedron with $n + 2$ faces. For $L$ large this solid has very low probability of landing on either of its two ends, whereas for very small $L$ it has a high probability of landing on one of them. Therefore by continuity there is some value of $L$ for which it has the same probability of landing on any one of its $n + 2$ faces. When $n$ is odd this yields a fair die with an odd number of faces.

The problem of characterizing all fair dice, not just those which are fair by symmetry or by continuity, is still unsolved.

It is a pleasure to thank Hans Samelson, Halsey Royden, Edward Gilbert, and Branko Grünbaum for interesting discussions of this topic.

REFERENCES

1. Fundamentals of Mathematics, Volume II, Geometry, H. Behnke, F. Bachman, K. Fladt, and H. Kunle, editors; translated by S. H. Gould, MIT Press, Cambridge, 1974.
2. Peter Pearce, Structure in Nature is a Strategy for Design, MIT Press, Cambridge, 1978.
3. Alan Holden, Shapes, Space and Symmetry, Columbia University Press, New York, 1971.
4. Branko Grünbaum and G. C. Shephard, Spherical tilings with transitivity properties, in The Geometric Vein: The Coxeter Festschrift, C. Davis, B. Grünbaum and F. Shenk editors, Springer-Verlag, New York, 1982.
5. Branko Grünbaum, On polyhedra in $E^3$ having all faces congruent, *Bull. Research Council Israel*, 8F (1960) 215–218.
6. Joseph B. Keller, The probability of heads, *Amer. Math. Monthly*, 93 (1986 ) 191–197.

## A Remark on Euclid's Proof of the Infinitude of Primes

JOHN B. COSGRAVE
*Mathematics Department, Carysfort College, Blackrock, Co. Dublin, Ireland*

IN MEMORY OF NIKHIL BANERJEE

Every student of elementary number theory knows Euclid's proof that there are infinitely many prime numbers: for if $p_1, \dots, p_n$ are the first $n$ primes and $N$ is their product, then *every* prime dividing $(N + 1)$ is larger than $p_n$. If, however, we let $N_i^{(n)}$ be the product of the first $n$ primes with the prime $p_i$ $(1 \leqslant i \leqslant n)$ excluded, it does *not* follow that *every* prime divisor of $(N_i^{(n)} + 1)$ is larger than $p_n$, as the prime $p_i$ is now a possible divisor, for example:

(a) $N_2^{(5)} + 1 = 2 \cdot 5 \cdot 7 \cdot 11 + 1 = 3 \cdot 257$, is divisible by the primes 3 and 257, $N_4^{(6)} + 1 = 2 \cdot 3 \cdot 5 \cdot 11 \cdot 13 + 1 = 7 \cdot 613$, divisible by the primes 7 and 613.

(b) $N_1^{(3)} + 1 = 3 \cdot 5 + 1 = 2^4$, is divisible only by the prime 2,
$N_1^{(2)} + 1 = 3 + 1 = 2^2$, also divisible only by the prime 2,
$N_2^{(2)} + 1 = 2 + 1 = 3$.

I used to mention these to my students to emphasize the success of Euclid's argument, particularly those in (b), to show that $(N_i^{(n)} + 1)$ need not be divisible by *any* prime larger than $p_n$. Experimentation easily leads to the discovery of more examples of type (a), but, as I will show in this note, there are no more of type (b).

If $(N_i^{(n)} + 1)$ is not divisible by a prime larger than $p_n$, then it must be a power of the excluded prime $p_i$, and so equal $p_i^m$, for some integer $m \geqslant 1$.

THEOREM. *$N_i^{(n)} + 1 = p_i^m$ is impossible for $n \geqslant 4$, $m$ an integer, $m \geqslant 1$.*

REMARK. $N_1^{(5)} + 1 = 3 \cdot 5 \cdot 7 \cdot 11 + 1 = 34^2$, so $(N_i^{(n)} + 1)$ can be a proper power, even when $n \geqslant 4$.

We require some lemmas.

LEMMA 1. *$p_n < p_1 \cdots p_{n-1} + 1$ for $n \geqslant 3$.*

*Proof.* This is well known. One has $p_{n+1} \leqslant (N + 1)$ for $n \geqslant 1$ (from Euclid)—which is to say that $p_n \leqslant (p_1 \cdots p_{n-1} + 1)$ for $n \geqslant 2$. But if one considers $(N - 1)$, with $n \geqslant 2$, it is divisible by a prime which must be larger than $p_n$, and the inequality $p_{n+1} \leqslant (N - 1) < (N + 1)$ follows. Thus $p_{n+1} < (p_1 \cdots p_n + 1)$ for $n \geqslant 2$, and so $p_n < (p_1 \cdots p_{n-1} + 1)$ for $n \geqslant 3$.

LEMMA 2. *If $p$ is prime and $x^p \equiv 1(p)$, then $x^p \equiv 1(p^2)$.*

*Proof.* Since $x^p \equiv 1(p)$, then $x \not\equiv 0(p)$, and so $x^{p-1} \equiv 1(p)$ by Fermat's little theorem. Thus $x^p - x^{p-1} \equiv 1 - 1 \equiv 0(p)$, and so $x^{p-1}(x - 1) \equiv 0(p)$, and $x \equiv 1(p)$. But $x^p - 1 = (x - 1)(x^{p-1} + x^{p-2} + \cdots + x + 1)$, and $(x^{p-1} + \cdots + 1) \equiv (1 + \cdots + 1) \equiv 0(p)$, and $(x - 1) \equiv 0(p)$, and so $(x^p - 1) \equiv 0(p^2)$.

LEMMA 3. *If $p$ and $q$ are primes and $x^p \equiv 1(q)$, then $x \equiv 1(q)$ or $q \equiv 1(p)$.*

*Proof.* Let $r = \mathrm{ord}_q x$, then $r|p$ and so $r = 1$ or $r = p$. If $r = 1$, then $x \equiv 1(q)$; and if $r = p$ then, since $x^{q-1} \equiv 1(q)$ by Fermat's theorem (as $x \not\equiv 0(q)$), $r|q - 1$, and so $q \equiv 1(p)$.

LEMMA 4. *If $p$ and $q$ are primes with $p^s \equiv 1(q)$, $s = p^k$ ($k$ an integer, $k \geqslant 1$) and $q < p$, then $p \equiv 1(q)$.*

*Proof.* If $k = 1$ then $p^p \equiv 1(q)$, and by Lemma 3 we have $p \equiv 1(q)$ or $q \equiv 1(p)$, and so $p \equiv 1(q)$ as $q \equiv 1(p)$ is ruled out when $q < p$. Since $p^{k+1} = p \cdot p^k$, a simple induction argument completes the proof.

LEMMA 5. *The Theorem is true for $i = 1$ and $i = 2$.*

*Proof.* Let $i = 1$ and suppose $N_1^{(n)} + 1 = 2^m$ for some integers $m$ and $n$ with $n \geqslant 4$. Then $2^m \equiv 1(3)$ and $2^m \equiv 1(7)$; thus $2|m$ since $\mathrm{ord}_3 2 = 2$, also $3|m$ since $\mathrm{ord}_7 2 = 3$, and so $6|m$. Thus $2^m = 2^{6M} = 64^M \equiv 1(9)$, as $64 \equiv 1(9)$. But then $3^2|2^m - 1$, whereas $N_1^{(n)}$ is square-free. So the Theorem is true for $i = 1$.

Now let $i = 2$, and suppose $N_2^{(n)} + 1 = 3^m$ for some integers $m$ and $n$ with $n \geqslant 4$. Then $3^m \equiv 1(5)$, and so $4|m$ since $\mathrm{ord}_5 3 = 4$. Thus $3^m = 3^{4M} = 81^M \equiv 1(16)$, as $81 \equiv 1(16)$. But then $2^4|3^m - 1$, whereas $N_2^{(n)}$ is square-free, and so the Theorem is true for $i = 2$.

*Proof of Theorem.* Suppose that $N_i^{(n)} + 1 = p_i^m$ for some integers $m$, $n$ and $i$ with $n \geqslant 4$ and $1 \leqslant i \leqslant n$; by Lemma 5 we may suppose that $3 \leqslant i \leqslant n$. Clearly $m \geqslant 1$, and $m = 1$ is impossible: for if $m = 1$ and $i < n$ then $p_i < p_n < N_i^{(n)}$ and $N_i^{(n)} + 1 = p_i$ is false; and if $m = 1$ and $i = n$, then $N_i^{(n)} + 1 = p_i$ gives $p_n = p_1 \cdots p_{n-1} + 1$, which is false for $n \geqslant 3$ by Lemma 1. Thus $m > 1$ and so $m$ is divisible by some prime $p$; our aim now is to determine the value of this prime $p$—it turns out that it must equal $p_i$, and it is that which enables us to make progress.

There are three possibilities: (a) $p = p_i$; (b) $p = p_j$, $1 \leqslant j \leqslant n$, $j \neq i$; (c) $p > p_n$.

If (b) holds then $p_i^m = x^p \equiv 1(p_j)$, and so $x^p \equiv 1(p)$; but then by Lemma 2 $x^p \equiv 1(p^2)$, and so $(p_i^m - 1)$ is divisible by $p^2$, whereas $N_i^{(n)}$ is square-free. Thus (b) cannot hold.

Suppose (c) holds. Now $N_i^{(n)} = p_i^m - 1 = x^p - 1 \equiv 0(p_j)$, $1 \leqslant j \leqslant n$, $j \neq i$, and so by Lemma 3 we have $x \equiv 1(p_j)$ or $p_j \equiv 1(p)$. But $p_j \equiv 1(p)$ doesn't hold when $p > p_n$, and so we have that $x \equiv 1(p_j)$ for *every* $j$ with $1 \leqslant j \leqslant n$ and $j \neq i$. Hence $x \equiv 1(N_i^{(n)})$, and so $(x - 1) \geqslant N_i^{(n)}$. But $N_i^{(n)} = x^p - 1 > x - 1 \geqslant N_i^{(n)}$, is impossible, and so (c) cannot hold.

Thus (a) holds and so the only prime divisor of $m$ is $p_i$, and so $m = p_i^k$ for some integer $k \geqslant 1$. Hence, applying Lemma 4 to $p_i^m \equiv 1(p_j)$, with $j$ restricted to the range $1 \leqslant j \leqslant (i - 1)$, we obtain $p_i \equiv 1(p_j)$, and thus $p_i \equiv 1(p_1 \cdots p_{i-1})$. But then $p_i \geqslant (p_1 \cdots p_{i-1}) + 1$, which, by Lemma 1, is impossible for $i \geqslant 3$. Thus the Theorem is proved.

COROLLARY. $p_{2n} \leqslant (p_2 \cdot p_3 \cdots p_n) - 2$ *for* $n \geqslant 3$.

*Proof.* Equality holds when $n = 3$. Now suppose that $n \geqslant 4$, then each $(N_i^{(n)} + 1)$, $1 \leqslant i \leqslant n$, must be divisible by some prime $q_i$ with $q_i > p_n$, and it is easy to see that the primes $q_1, \dots, q_n$ are distinct: for suppose $q_i = q_j = q$ for some $i$ and $j$ with $1 \leqslant i, j \leqslant n$ and $i \neq j$, then $q | (N_i^{(n)} + 1)$ and $q | (N_j^{(n)} + 1)$, and so $q | (N_i^{(n)} - N_j^{(n)})$. But $N_i^{(n)} - N_j^{(n)} = (p_1 \cdots p_n) \cdot (p_j - p_i)/p_i p_j$, and so $q | (p_j - p_i)$, which is impossible, since $1 \leqslant |p_j - p_i| < p_n$.

Thus $q_1, \dots, q_n$ are $n$ distinct primes, each larger than $p_n$, and so we have: $p_{2n} \leqslant \max_{1 \leqslant i \leqslant n} q_i \leqslant \max_{1 \leqslant i \leqslant n} (N_i^{(n)} + 1) = (p_2 \cdot p_3 \cdots p_n) + 1$, for $n \geqslant 4$, and it follows that $p_{2n} \leqslant (p_2 \cdot p_3 \cdots p_n) - 2$, since the integers $(p_2 \cdot p_3 \cdots p_n) \pm 1$ (being even and larger than 2) and $(p_2 \cdot p_3 \cdots p_n)$ are composite.

*Note.* The Theorem is trivial if $p_i \equiv 1(4)$ (but not for $p_i \equiv 3(4)$) as $N_i^{(n)} + 1 = p_i^m$ is clearly impossible, since $4 | (p_i^m - 1)$ but $N_i^{(n)}$ is square-free.

Finally, the proof given here has been modified to obtain:

THEOREM. $N_i^{(n)} - 1 = p_i^m$ *is impossible for* $n \geqslant 4$. (*Note:* $N_2^{(3)} - 1 = 2 \cdot 5 - 1 = 3^2$).

COROLLARY. $p_{3n-1} \leqslant (p_2 \cdot p_3 \cdots p_n) - 2$ *for* $n \geqslant 4$. (*Yes, that is* $(3n - 1)$, *and not* $3n$).

*Proofs.* Available upon request.

# An Alternate Proof of the Continuity of the Roots of a Polynomial

FELIPE CUCKER AND ANTONIO GONZALEZ CORBALAN
*Depto. de Matematicas, Facultad de Ciencias, Universidad de Santander, Spain*

**Introduction.** Let $P(X) = X^n + a_1 X^{n-1} + \cdots + a_n$ be a monic complex polynomial. The Fundamental Theorem of Algebra states that there are $\xi_1, \ldots, \xi_n$ belonging to $\mathbb{C}$ such that $P(X) = \prod_{1 \leqslant i \leqslant n}(X - \xi_i)$.

Of course, if we vary the coefficients $a_1, \ldots, a_n$, the root system $\xi_1, \ldots, \xi_n$ will also vary. The continuity of this variation is a well known theorem that can be stated more precisely in the following way.

THEOREM 1. *Let $P(X) = X^n + a_1 X^{n-1} + \cdots + a_n$ be a monic complex polynomial, and let $\xi_1, \ldots, \xi_n$ be its roots. Given $\varepsilon \in \mathbb{R}$, $\varepsilon > 0$, there is a $\delta \in \mathbb{R}$, $\delta > 0$, such that for every monic polynomial $Q(X) = X^n + b_1 X^{n-1} + \cdots + b_n$, if $|b_j - a_j| < \delta$ for $1 \leqslant j \leqslant n$, then there are $\zeta_1, \ldots, \zeta_n$ belonging to $\mathbb{C}$ such that $Q(X) = \prod_{1 \leqslant i \leqslant n}(X - \zeta_i)$ and $|\zeta_j - \xi_j| < \varepsilon$ for $1 \leqslant j \leqslant n$.*

Proofs of this theorem can be found in many places, for instance [**1**], [**2**], or [**3**]. The usual proofs make use of Rouché's theorem. Our goal is to give an alternate proof which does not. We begin by stating some facts about symmetric functions and projective spaces.

**An equivalent statement.** We can consider $P(X)$ as a point in $\mathbb{C}^n$, namely the point $(a_1, \ldots, a_n)$ of its coefficients. However, we cannot do the same thing with the root system $\xi_1, \ldots, \xi_n$ because it is unordered; permutation of the $\xi_i$'s gives us the same root system. In order to resolve that we introduce the following equivalence relation in $\mathbb{C}^n$: $(\xi_1, \ldots, \xi_n) \simeq (\zeta_1, \ldots, \zeta_n)$ if and only if there is a permutation $\sigma$ in the symmetric group $S_n$ such that $\zeta_j = \xi_{\sigma(j)}$ for $1 \leqslant j \leqslant n$. This is clearly an equivalence relation in $\mathbb{C}^n$, and we will denote by $\mathbb{C}^n_{\text{sym}}$ the quotient given by this relation.

Now, we have a map $\tau: \mathbb{C}^n \to \mathbb{C}^n_{\text{sym}}$ sending the polynomial $P(X)$ into its root system. This is a bijective map because two monic polynomials coincide if and only if they have the same root system, and every $n$-tuple in $\mathbb{C}^n$ determines a monic polynomial having it as root system. Moreover, the inverse map is well known and is given by

$$\sigma: \mathbb{C}^n_{\text{sym}} \to \mathbb{C}^n$$

$$[\xi_1, \ldots, \xi_n] \to (\sigma_1(\xi_1, \ldots, \xi_n), \ldots, \sigma_n(\xi_1, \ldots, \xi_n)),$$

where $\sigma_1, \ldots, \sigma_n$ are the elementary symmetric functions in $n$ variables.

We can now restate Theorem 1. We endow $\mathbb{C}^n_{\text{sym}}$ with the quotient topology: a subset $U \subset \mathbb{C}^n_{\text{sym}}$ is open if and only if its pre-image $U$ in $\mathbb{C}^n$ is open. Then Theorem 1 becomes:

THEOREM 1$'$. *$\sigma$ is a homeomorphism.*

We note that $\sigma$ is continuous because it is induced by a polynomial, and the continuity of $\sigma^{-1}$ $(= \tau)$ is just the continuity of the roots as functions of the coefficients.

**Extension to projective space.** The idea of our proof is to prove a corresponding result for an extension of $\sigma$ to a bijection $\sigma^*$ defined on a compact space containing $\mathbb{C}^n_{\text{sym}}$. To do that, we consider projective spaces over $\mathbb{C}$. This idea is hidden in the following informal remark at the end of the proof of Theorem 1 in **[2]**: "Let us suppose that the coefficient of the highest power of $X$ is not unity; we can now write the equation as

$$a_0X^n + a_1X^{n-1} + \cdots + a_n = 0.$$

If in this equation we make $a_0$ tend to 0, one of these roots becomes greater than every possible limit while the others have as limits the roots of the equation of degree $n-1$

$$a_1X^{n-1} + a_2X^{n-2} + \cdots + a_n = 0.$$

Moreover if $a_1$ is zero or $a_1$ and $a_2$ are both zero, two or three roots become infinite."

The $n$-dimensional projective space over $\mathbb{C}$ is the set $\mathbb{P}^n(\mathbb{C})$ of equivalence classes of points $(x_0,\dots,x_n)$ in $\mathbb{C}^{n+1} - \{(0,\dots,0)\}$ under the relation: $(x_0,\dots,x_n) \simeq (y_0,\dots,y_n)$ if and only if there is a nonzero $\lambda \in \mathbb{C}$ such that $y_j = \lambda x_j$ for $0 \leqslant j \leqslant n$. As every equivalence class has a point in the unit sphere $S^n$, we can see $\mathbb{P}^n(\mathbb{C})$ as a quotient of $S^n$. With the quotient topology, $\mathbb{P}^n(\mathbb{C})$ is then a compact space because $S^n$ is and a quotient of a compact space is compact.

The point in $\mathbb{P}^n(\mathbb{C})$ whose equivalence class contains $(x_0,\dots,x_n)$ will be denoted by $\langle x_0,\dots,x_n\rangle$. Now we consider the injection

$$\mathbb{C}^n \to \mathbb{P}^n(\mathbb{C})$$
$$(x_1,\dots,x_n) \to \langle 1, x_1,\dots,x_n\rangle,$$

which is a homeomorphism from $\mathbb{C}^n$ onto its image in $\mathbb{P}^n(\mathbb{C})$, the set of points with nonzero first coordinate.

The correspondence between monic polynomials of degree $n$ and points in $\mathbb{C}^n$ can now be extended. Two polynomials $P(X) = a_0X^n + a_1X^{n-1} + \cdots + a_n$ and $Q(X) = b_0X^n + b_1X^{n-1} + \cdots + b_n$ are said to belong to the same *homotesy* class when there is a $\lambda \in \mathbb{C}$, $\lambda \neq 0$, such that $b_j = \lambda a_j$ for $0 \leqslant j \leqslant n$. This is equivalent to saying that $P(X)$ and $Q(X)$ have the same degree and the same root system. We associate the point $\langle a_0,\dots,a_n\rangle$ to the homotesy class of the polynomial $a_0X^n + a_1X^{n-1} + \cdots + a_n$. A point $\langle a_0,\dots,a_n\rangle$ in $\mathbb{P}^n(\mathbb{C})$ then belongs to $\mathbb{C}^n$ if and only if the elements of the corresponding homotesy class are polynomials of degree $n$.

In preparation for the definition of $\sigma^*$, we define extensions of the elementary symmetric functions,

$$\sigma_i^h\colon \left((\mathbb{C}^2)\right)^n \to \mathbb{C},$$

by

$$\sigma_i^h((z_1,x_1),\dots,(z_n,x_n)) = z_1 \cdots z_n\sigma_i(x_1/z_1,\dots,x_n/z_n).$$

It is easy to check that

$$\sigma_i^h((z_1,x_1),\dots,(z_n,x_n)) = (-1)^i\sum x_{k_1}\cdots x_{k_i}z_{k_{i+1}}\cdots z_{k_n},$$

where $\{k_1,\dots,k_i\}$ runs over all possible choices of $i$ elements in $\{1,\dots,n\}$, and $\{k_{i+1},\dots,k_n\}$ are the remaining elements. Then $\sigma_i^h$ is a homogeneous polynomial

in the variables $z_1,\dots,z_n,x_1,\dots,x_n$ of total degree $n$ satisfying

$$\sigma_i^h((\lambda_1 z_1,\lambda_1 x_1),\dots,(\lambda_n z_n,\lambda_n x_n)) = \lambda_1\cdots\lambda_n\sigma_i^h((z_1,x_1),\dots,(z_n,x_n)).$$

This last remark allows us to define

$$\sigma^h\colon (\mathbb{P}^1(\mathbb{C}))^n \to \mathbb{P}^n(\mathbb{C})$$

by

$$\sigma^h(\langle z_1,x_1\rangle,\dots,\langle z_n,x_n\rangle) = \langle z_1\cdots z_n,\dots,\sigma_i^h(\langle z_1,x_1\rangle,\dots,\langle z_n,x_n\rangle),\dots\rangle.$$

As the $\sigma_i^h$ are symmetric, $\sigma^h$ induces a function defined on $(\mathbb{P}^1(\mathbb{C}))^n_{\text{sym}}$, the quotient of $(\mathbb{P}^1(\mathbb{C}))^n$ by the action of $S_n$, which we denote

$$\sigma^*\colon (\mathbb{P}^1(\mathbb{C}))^n_{\text{sym}} \to \mathbb{P}^n(\mathbb{C}).$$

**The proof that $\sigma^*$ is a homeomorphism.** We begin by explaining the meaning of $\sigma^*$.

LEMMA. *Let $\langle 1,\xi_1\rangle,\dots,\langle 1,\xi_r\rangle,\langle 0,1\rangle,\dots,\langle 0,1\rangle$ be $n$ points in $\mathbb{P}^1(\mathbb{C})$, the first $r$ of which belong to $\mathbb{C}\subset\mathbb{P}^1(\mathbb{C})$. If*

$$\sigma^*([\langle 1,\xi_1\rangle,\dots,\langle 1,\xi_r\rangle,\langle 0,1\rangle,\dots,\langle 0,1\rangle]) = \langle a_0\cdots a_n\rangle,$$

*then $a_0=\cdots=a_{n-r-1}=0$, $a_{n-r}\neq 0$, and the polynomial $a_{n-r}X^r+\cdots+a_{n-1}X+a_n$ has $(\xi_1,\dots,\xi_r)$ as root system.*

*Proof.* It is easy to check that

$$\sigma_i^h((1,X_1),\dots,(1,X_r),(0,1),\dots,(0,1)) = \begin{cases} 0, & \text{if } 1\leqslant i\leqslant n-r-1;\\ 1, & \text{if } i=n-r;\\ \sigma_{i-(n-r)}(X_1,\dots,X_r), & \text{if } n-r+1\leqslant i\leqslant n;\end{cases}$$

where $\sigma_0^h = z_1\cdots z_n$.

The result follows immediately.

COROLLARY. *The map $\sigma^*$ is bijective, and its restriction to $\mathbb{C}^n_{\text{sym}}$ gives us the bijection $\sigma\colon\mathbb{C}^n_{\text{sym}}\to\mathbb{C}^n$.*

THEOREM 2. *$\sigma^*\colon(\mathbb{P}^1(\mathbb{C}))^n_{\text{sym}}\to\mathbb{P}^n(\mathbb{C})$ is a homeomorphism for the quotient topology in $(\mathbb{P}^1(\mathbb{C}))^n_{\text{sym}}$.*

*Proof.* Since $\mathbb{P}^1(\mathbb{C})$ is a compact space, the same is true for $(\mathbb{P}^1(\mathbb{C}))^n$ by Tychonoff's theorem, and then $(\mathbb{P}^1(\mathbb{C}))^n_{\text{sym}}$ is also a compact space. Now, $\sigma^*$ is continuous because it is induced by a polynomial, and a continuous bijective function from a compact space onto a Hausdorff space is a homeomorphism.

Theorem 1′ follows from this last theorem and the corollary, and with it Theorem 1 also follows. But we can deduce from Theorem 2 a stronger version of Theorem 1 giving precise meaning to Weber's remarks quoted above. It suffices to note that $\mathbb{P}^1(\mathbb{C})=\mathbb{C}\cup\{\langle 0,1\rangle\}$, and that a basis of neighborhoods of this last point is given by a family of sets of the form $\{\langle z,x\rangle\in\mathbb{P}^1(\mathbb{C})\,|\,z=0 \text{ or } |x/z|>K\}$ with $K\in\mathbb{R}$, $K>0$. Using this fact we get, by translating Theorem 2 into "epsilon-delta"

language, the following:

THEOREM 3. *Let* $P(X) = a_0X^n + a_1X^{n-1} + \cdots + a_n$ *be a nonzero complex polynomial of degree* $k \leqslant n$. *Let* $\xi_1, \ldots, \xi_r$ *be its roots in* $\mathbb{C}$, *with multiplicities* $m_1, \ldots, m_r$ *respectively, and let* $U_1, \ldots, U_r$ *be disjoint disks centered at* $\xi_1, \ldots, \xi_r$ *with radii* $\varepsilon$ *and contained in the open disk centered at* $0$ *with radius* $1/\varepsilon$. *Then there is a* $\delta \in \mathbb{R}$, $\delta > 0$, *such that, if* $|b_j - a_j| < \delta$ *for every* $0 \leqslant j \leqslant n$, *then the polynomial* $Q(X) = b_0X^n + b_1X^{n-1} + \cdots + b_n$ *has* $m_i$ *roots (counted with multiplicity) in each* $U_i$ *and* $\deg(Q(X)) - k$ *roots with absolute value greater than* $1/\varepsilon$.

**Some final remarks.** The different proofs of Theorem 1 also give different additional information. For instance, the proof in [**1**] contains numerical bounds leading to practical calculation, and the proof in [**3**] is done for arbitrary algebraically closed fields. We think the interest of our proof is that it shows the advantage of working in complex projective spaces, which leads to a proof using only basic facts from general topology. It may also be the shortest proof available.

REFERENCES

1. A. Ostrowski, Solution of Equations in Euclidean and Banach Spaces, Academic Press, 1973.
2. H. Weber, Traité d'algebre Supérieure, Gauthiers-Villars, 1898.
3. H. Whitney, Complex Analytic Manifolds, Addison-Wesley, 1972.

Partially supported DGCYT "Geometria y Algebra Reales"

EDITED BY MELVIN HENRIKSEN AND STAN WAGON

# Taylor's Theorem Using the Generalized Riemann Integral

H. B. THOMPSON
*Department of Mathematics, University of Queensland, St. Lucia, Queensland, Australia 4067*

**1. Introduction.** It is well known and often emphasized in texts (see **[16]**) that the Lagrange and Cauchy remainders are valid under the mere assumption of the existence of the $(n+1)$st derivative, whereas the integral form of the remainder requires some additional hypotheses, for example, integrability of the $(n+1)$st derivative (see **[17**, p. 281]). This additional hypothesis is necessary if integrals such as the Riemann or the Lebesgue integrals are used. In this note we show that when the generalized Riemann integral is used, no additional hypotheses are needed. Indeed, in this case the integral remainder is valid if the $(n+1)$st derivative exists except possibly at a countable number of points, and the $n$th derivative is continuous on the closed interval. Moreover, the Lagrange and Cauchy forms of the remainder can be deduced from the integral form in the special case where the $(n+1)$st derivative exists everywhere except possibly at the end points of the interval.

The generalized Riemann integral is variously known as the Kurzweil, the Riemann complete, and the gauge integral. It is also equivalent to the Perron integral, the descriptive $D^*$-integral of Luzin, and the restricted total integral (also called the $T^*$-integral) of Denjoy. Newton introduced integration as antidifferentiation. Between 1912 and 1915, Denjoy [3], Luzin **[10]**, and Perron **[15]**, realizing that the Lebesgue and Newton integrals did not properly contain one another, gave new definitions of the integral to encompass both the Newton and Lebesgue integrals. The equivalence of the Denjoy and Luzin integrals is not difficult to prove while the equivalence of these to the Perron integral is due to Hake **[4]**, Looman **[9]**, and Aleksandrov **[1]**. Kurzweil **[7]** introduced his integral for application to ordinary differential equations, and showed that it is equivalent to the Perron integral. Henstock **[5]** independently introduced this integral and developed its properties (see, e.g., **[6]**).

In the case $n = 0$, the integral remainder form of Taylor's theorem is just the fundamental theorem of calculus. Thus it is not surprising that using the generalized Riemann integral this form of Taylor's theorem should hold under such weak differentiability assumptions.

It is a pleasure to acknowledge the very useful comments made by R. Výborný and the referee.

**2. Preliminaries.** For a detailed discussion of the generalized Riemann integral see **[6, 8, 12, 13, 14]**. Our proof is a minor variation on the usual proof for the integral remainder in the case where all derivatives are continuous; so, we need the appropriate form of the fundamental theorem of calculus and its associated theorem on integration by parts. A nice presentation of this by Swartz and Thomson **[18]** can be found in this MONTHLY (see also **[14]**). For the sake of clarity we will give a brief outline of the definitions and basic properties of the generalized integral.

By a tagged partition $\mathbf{T}$ of $[a, b]$ we mean a set $\{x_0, x_1, \ldots, x_n; t_1, t_2, \ldots, t_n\}$ satisfying

$$a = x_0 \leqslant t_1 \leqslant x_1 \leqslant t_2 \leqslant x_2 \leqslant \cdots \leqslant x_n = b$$

for some $n > 0$. A positive function $\delta$: $[a, b] \to \mathbb{R}^+$ is called a gauge on $[a, b]$. Let $\delta$ be a gauge on $[a, b]$; then the partition $\mathbf{T}$ is said to be $\delta$-fine if

$$[x_{i-1}, x_i] \subseteq (t_i - \delta(t_i), t_i + \delta(t_i))$$

for $i = 1, 2, \ldots, n$.

Using bisection and the nested interval theorem it is easy to prove that for every gauge $\delta$ on $[a, b]$ there exists a $\delta$-fine partition of $[a, b]$.

Definition. Let $f$:$[a, b] \to \mathbb{R}$. Then $I$ is said to be the generalized Riemann integral of $f$ on $[a, b]$ (denoted by $\int_a^b f(t)\,dt$) if, given $\epsilon > 0$, there exists a gauge $\delta$ on $[a, b]$ such that

$$\left| \sum_{i=0}^{n} f(t_i)(x_i - x_{i-1}) - I \right| < \epsilon,$$

whenever the partition $\mathbf{T}$ is $\delta$-fine. We call $f$ integrable on $[a, b]$ if its generalized Riemann integral exists.

We are ready to state the fundamental theorem and its associated theorem on integration by parts.

Definition. (See [**14**].) Let $f$:$[a, b] \to \mathbb{R}$ be given. A function $F$: $[a, b] \to \mathbb{R}$ is a *primitive* of $f$ on $[a, b]$ provided $F$ is continuous on $[a, b]$ and $F'(x) = f(x)$ for all $x$ in $[a, b]$, except possibly at a finite or countably infinite set of values of $x$.

The Fundamental Theorem. ([**18**, Thm. 5]) *If $f$ has a primitive $F$ on $[a, b]$, then $f$ is integrable and*

$$\int_a^b f(t)\,dt = F(b) - F(a). \tag{1}$$

*Remark*. To guarantee the validity of (1) and of the integral form of Taylor's theorem in the case of the Riemann or Lebesgue integrals, additional assumptions such as the integrability of $f$ are required. In particular, there is a function $F$ having a bounded derivative everywhere on $[a, b]$ but such that $f = F'$ is not Riemann integrable on $[a, b]$. Also, the function $F$ defined by $F(x) = x^2 \sin 1/x^2$, for $x \neq 0$, $F(0) = 0$ is differentiable everywhere but $f = F'$ is not Lebesgue integrable on $[a, b]$ if $a \neq b$ and $0 \in [a, b]$.

As an immediate consequence of the fundamental theorem, one obtains the following.

Integration by Parts. (See [**14**, p. 50].) *If $g$ and $h$ have primitives $G$ and $H$, respectively, on $[a, b]$, then $gH$ is integrable if and only if $Gh$ is integrable. Moreover*

$$\int_a^b g(t)H(t)\,dt = G(b)H(b) - G(a)H(a) - \int_a^b G(t)h(t)\,dt. \tag{2}$$

*Remark*. The integrability of $gH$ and hence of $Gh$ is necessary for (2) to hold as can be seen by setting

$$F(x) = x^2 \sin x^{-4}, \qquad G(x) = x^2 \cos x^{-4} \quad \text{for} \quad x \neq 0 \quad \text{and} \quad F(0) = 0 = G(0).$$

See [**11**, Ex. 13]. In our case $g$ will be continuous on $[a, b]$ so $gH$ will be integrable on $[a, b]$.

Some other advantages of using this definition of the integral, especially for teaching, are as follows (see [**20**]):

(a) The definition is based on Riemann sums so it is easy to grasp intuitively.
(b) Most of the elementary theory is as easy as it is for the Riemann integral. Also the monotone and dominated convergence theorems can be proved fairly early in the development of the theory (see [**6**]).
(c) Improper integrals are automatically included. In particular, if $f$ is generalized Riemann integrable on $[a, c]$ for every $c$ with $a \leqslant c < b$, then $\int_a^b f(t)\,dt$ exists and is equal to $\lim_{c \to b} \int_a^c f(t)\,dt$ iff this limit exists (see [**14**]).
(d) There is a simple direct proof of the integral formulas for arc length in $\mathbb{R}^k$ assuming only the differentiability of the parameterization (see [**20**]).
(e) For functions of one variable a Lebesgue integrable function is generalized Riemann integrable with the same value for its integral. For nonnegative functions the generalized Riemann and Lebesgue integrals coincide.

## 3. The Main Result.

TAYLOR'S THEOREM. *Let $f, f^{(1)}, \ldots, f^{(n)}$ be continuous on $[a, b]$ and suppose that $f^{(n+1)}$ exists on $[a, b]$, except possibly at a countable number of points; then,*

$$f(b) = f(a) + f^{(1)}(a)(b-a) + \cdots + f^{(n)}(a)(b-a)^n/n! + R_{n,a}(b), \tag{3}$$

*where*

$$R_{n,a}(b) = \int_a^b f^{(n+1)}(t)(b-t)^n/n!\,dt. \tag{4}$$

(That $R_{n,a}(b)$ is well defined under the hypotheses is part of the conclusion.)

*Proof.* We use induction. For $n = 0$ the result is the fundamental theorem. Assume that (3) holds for $n = k$, where the remainder given by (4) exists; it suffices to show that the remainder $R_{k+1,a}(b)$ exists and

$$R_{k,a}(b) = f^{(k+1)}(a)(b-a)^{k+1}/(k+1)! + R_{k+1,a}(b);$$

that is,

$$\begin{aligned}\int_a^b f^{(k+1)}(t)(b-t)^k/k!\,dt &= f^{(k+1)}(a)(b-a)^{k+1}/(k+1)! \\ &\quad + \int_a^b f^{(k+2)}(t)(b-t)^{k+1}/(k+1)!\,dt,\end{aligned} \tag{5}$$

where both sides exist. But this follows from (2) with $H(t) = f^{(k+1)}(t)$ and $G(t) = (b-t)^{k+1}/(k+1)!$.

*Remark.* The above proof relies only on the fundamental theorem so that the extensions to the complex and to the vector-valued cases are immediate.

## 4. Other forms of the remainder.

We briefly indicate how the other forms of the remainder follow from the integral form. See [**19**] for more details. The following notation and results will be helpful.

We say that $f:(a, b) \to \mathbb{R}$ has the intermediate value property on $(a, b)$ if given $u, v$ in $(a, b)$ and $y$ such that $f(u) < y < f(v)$ there exists $s$, strictly between $u$ and $v$, such that $y = f(s)$. By Darboux's Theorem (see [**17**, p. 186]), derivatives have the intermediate value property.

The following mean value theorems can, with a little care, be proved along the lines used for their Riemann integral analogues (see, e.g. [**19**, p. 4]).

THE FIRST MEAN VALUE THEOREM. *Let $f$ be generalized Riemann integrable on $[a, b]$ and have the intermediate value property on $(a, b)$. Then*

$$\int_a^b f(t)\, dt = f(s)(b-a)$$

*for some $s$, $a < s < b$.*

THE GENERALIZED MEAN VALUE THEOREM. *Let $f$ have the intermediate value property and assume that $g$ does not change sign on $(a, b)$. If $g$ and $fg$ are generalized Riemann integrable on $[a, b]$, then*

$$\int_a^b f(t) g(t)\, dt = f(s) \int_a^b g(t)\, dt,$$

*for some $s$, $a < s < b$.*

Bonnet's second mean value theorem and the generalized second mean value theorem also hold for the generalized Riemann integral.

TAYLOR'S THEOREM (Cauchy and Lagrange Remainders). *Let $f, f^{(1)}, \dots, f^{(n)}$ be continuous on $[a, b]$ and suppose that $f^{(n+1)}$ exists on $(a, b)$. Then*

$$f(b) = f(a) + f^{(1)}(a)(b-a) + \cdots + f^{(n)}(a)(b-a)^n/n! + R_{n,a}(b), \tag{6}$$

*where*

$$R_{n,a}(b) = f^{(n+1)}(s)(b-a)^{n+1}/(n+1)!, \tag{7}$$

*for some $s$, $a < s < b$ (Cauchy remainder), and*

$$R_{n,a}(b) = f^{(n+1)}(u)(b-u)^n(b-a)/n! \tag{8}$$

*some $u$, $a < u < b$ (Lagrange remainder).*

*Proof.* From (4) and (6), $R_{n,a}(b) = \int_a^b f^{(n+1)}(t)(b-t)^n/n!\, dt$. Let

$$g(t) = f(t) + \{(b-t)f^{(1)}(t) + (b-t)^2 f^{(2)}(t)/2! + \cdots + (b-t)^n f^{(n)}(t)/n!\}.$$

Thus $g'(t) = f^{(n+1)}(t)(b-t)^n/n!$ has the intermediate value property on $(a, b)$, by Darboux's theorem, so (8) holds by the first mean value theorem.

Now $f^{(n+1)}(t)$ has the intermediate value property on $(a, b)$ and $g(t) = (b-t)^n/n!$ does not change sign on $(a, b)$, so

$$\begin{aligned}\int_a^b f^{(n+1)}(t)(b-t)^n/n!\, dt &= f^{(n+1)}(u)\int_a^b (b-t)^n/n!\, dt \\ &= f^{(n+1)}(u)(b-a)^{(n+1)}/(n+1)!\end{aligned}$$

for some $u$, $a < u < b$, by the generalized first mean value theorem. Thus (7) holds.

Young's remainder (see [2, p. 80]) and various other remainders can also be derived from the integral remainder.

REFERENCES

1. A. Aleksandrov, Über die Äquivalenz des Perronschen und des Denjoyschen Integralbegriffes, *Math. Z.*, 20 (1924) 213–222.
2. J. C. Burkill, A First Course in Mathematical Analysis, Cambridge University Press, Cambridge, 1967.
3. A. Denjoy, Une extension de l'intégrale de M. Lebesgue, *C. R. Acad. Sci. Paris*, 154 (1912) 895–862.
4. H. Hake, Über de la Vallée Poussins Ober-und Unterfunktionen, *Math. Ann.*, 83 (1921) 119–142.
5. R. Henstock, Definitions of Riemann type of variational integrals, *Proc. London Math. Soc.*, **1** (1961) 402–418.
6. ______, A Riemann integral of Lebesgue power, *Canadian J. Math.*, 20 (1968) 79–87.
7. J. Kurzweil, Generalized ordinary differential equations and continuous dependence on a parameter, *Czechoslovak Math. J.*, 82 (1957) 418–449.
8. ______, On Fubini Theorem for general Perron integral, *Czechoslovak Math. J.*, 98 (1973) 286–297.
9. H. Looman, Über die Perronsche Integral Definition, *Math. Ann.*, 93 (1935) 153–156.
10. N. N. Luzin, Sur les propriétés de l'intégrale de M. Denjoy, *C. R. Acad. Sci. Paris*, 155 (1912) 1475–1478.
11. J. Mărík, Foundation of the theory of an integral in Euclidean space (translated into English by L. I. Trudzik, Dept. of Math., Univ. of Melbourne, Parkville, Victoria 3052, Australia), *Časopis Pěst. Mat.*, 77 (1952) 125–144.
12. J. Mawhin, L'introduction a l'Analyse, CABAY, Louvain-La-Neuve, 1984.
13. E. J. McShane, A Riemann-type Integral that Includes Lebesgue-Stieltjes, Bochner and Stochastic Integrals, Vol. 88, Memoirs of Amer. Math. Soc., 1969.
14. R. M. McLeod, The Generalized Riemann Integral, Carus Mathematical Monographs, Mathematical Association of America, 1980.
15. O. Perron, Über den Integralbegriff, Sitzber, *Heidelberg Akad. Wiss. Abt. A*, 16 (1914) 1–16.
16. M. Spivak, Calculus, W. A. Benjamin, New York, 1967.
17. K. R. Stromberg, An Introduction to Classical Real Analysis, Wadsworth, Belmont, California, 1981.
18. C. Swartz and B. S. Thomson, More on the fundamental theorem of calculus, this MONTHLY, 95 (1987) 644.
19. H. B. Thompson, Taylor's theorem with the integral remainder under very weak differentiability assumptions, *Austral. Math. Soc. Gazette*, 12 (1985) 1–6.
20. R. Výborný, Kurzweil's integral and arclength, *Austral. Math. Soc. Gazette*, 8 (1981) 19–22.

# The Effect of Prior Calculus Experience on "Introductory" College Calculus

MARTHA B. BURTON
*Mathematics Center, University of New Hampshire, Durham, NH 03824*

**Introduction.** This report concerns the performance of students in the introductory calculus course at the University of New Hampshire (UNH), especially with respect to their experience of calculus in high school. Although the failure rate in this course does not approach the 35% national failure rate in calculus [1], it is high enough to cause local concern.

One of the peculiar characteristics of a first-year university calculus course is the extent of its reliance on specific mathematics skills that students are presumed to

have mastered in high school. College-level courses in some fields have no exact counterparts in the high school curriculum and, hence, afford the student something of the luxury of a fresh start. Other courses might chiefly require a certain level of student maturity and a general fund of knowledge. But calculus courses, and for that matter first-year courses in some of the physical sciences, are unforgiving: they depend heavily and directly upon the student's actual skills from high school mathematics courses. This circumstance has naturally led to heightened concern about mathematics in the student transition from high school to college.

In recent years the advice given by the UNH Mathematics Department to high school mathematics departments has largely coincided with that of the CUPM panel on calculus articulation [**2**]. In particular, it has been urged that calculus taught in high schools should be a year course based on the Advanced Placement syllabus, and that where this is not feasible a sound preparation in algebra and trigonometry is preferable to a superficial treatment of the calculus.

With these concerns in mind, and as a result of many conversations with students, a single informational question about the high school calculus experience of our students was appended to the final examination of the first-semester calculus course in the fall of 1987.

The results of this simple survey appear to offer a warning footnote to the advice of the CUPM panel about calculus in high schools. It is certainly rational for high schools without the resources for a serious AP calculus course to refrain from substituting a year-long but watered-down calculus course for it. Nevertheless, from the individual student's perspective, a year of high school calculus of even this lesser quality seems to have afforded the student a noticeable advantage in the standard first-year course at UNH.

Conversely and worse, the lack of the year of high school calculus can seriously handicap the first-year university student, who must compete with a significant proportion of classmates for whom the subject is not new. Indeed, the competitive disadvantage of the student without high school calculus serves to illustrate another point made by the CUPM panel: that the number of high school students taking calculus has been increasing and that many students now enter college with a high school calculus background, not necessarily of the form recommended by the CUPM panel. In the present survey, less than a quarter of the students with high school calculus experience described it as an Advanced Placement course.

Students who have taken calculus in high school do not, of course, always find themselves in the first-semester calculus course at UNH. Some begin instead with the second semester of calculus, or in an Honors calculus course, or with differential equations. Nevertheless there is now a sizeable population of students with high school calculus experience in the standard first-semester calculus course—a course which does not presuppose calculus, but is intended for students with a good precalculus background.

**Survey results.** The 741 students who took the final examination for the first-semester calculus course in December 1987 were asked to describe their high school calculus experience as: (1) no previous calculus experience; (2) a brief introduction to calculus topics; (3) a full year of high school calculus; or (4) a full year of Advanced Placement high school calculus. The course letter-grade results seemed to

show two, not four, distinct categories;

| | A | B | C | D | F |
|---|---|---|---|---|---|
| Minimal (or no) previous calculus (370) | 9 | 53 | 99 | 94 | 115 |
| Full year of high school calculus (313) | 47 | 119 | 95 | 33 | 19 |

In this "introductory" course, there are now almost as many students who have had a previous year of calculus in high school as there are beginning students. Further, it appears that the students with the preparatory year of high school calculus, while doing well as a group, are not overqualified for the course: their letter-grades accumulate in the Bs and Cs, not in the As. The letter-grade performance of the students with minimal prior calculus experience, taken as a group, is very disappointing.

**Other factors: the calculus pretest.** Another factor known to predict success in the first-semester calculus course is an MAA test in algebra and trigonometry, used as a Calculus Pretest during the first week of the course. Passing the calculus pretest at the beginning of the term means that the student is excused from mandatory remediation, during the calculus course, in individually-prescribed topics in algebra and trigonometry. Not surprisingly, students who pass the pretest are much more likely to do well in the calculus course than those who don't. And the hazard that apparently attaches to the lack of prior calculus experience weighs less heavily upon students who are able to pass the calculus pretest:

| | | A | B | C | D | F |
|---|---|---|---|---|---|---|
| Minimal (or no) previous calculus: | need remediation | 0 | 9 | 37 | 31 | 80 |
| | passed pretest | 8 | 43 | 57 | 57 | 30 |
| Full year of high school calculus: | need remediation | 3 | 20 | 35 | 18 | 12 |
| | passed pretest | 42 | 96 | 59 | 13 | 6 |

The group at risk would seem to be students who had no prior calculus experience *and* who could not pass the calculus pretest. That group reached a 50% failure rate, and accounted for almost two-thirds of the failing grades in the survey group.

The second line of the table is disquieting, for it represents students who, with minimal or no prior calculus experience, still passed the algebra-trigonometry calculus pretest. These are the students who come closest to reflecting the stated prerequisites for the course: a sound precalculus high school background, including trigonometry. Still, their letter-grade results are disappointing.

**Other factors: SAT scores.** The average quantitative SAT score for the group with minimal or no high school calculus was 575 with a standard deviation of 67; for the group with one year of high school calculus the same scores showed an

average of 615 with a standard deviation of 64. The overlap in the groups is more striking than the difference in their average scores. It would be hard to justify any elitist distinction between a "calculus group" and a "noncalculus group" in high school on the basis of the resulting quantitative SAT scores. Quite to the contrary, it might well be expected that whatever enrichment in the high school mathematics curriculum a "calculus group" enjoyed would enhance its members' SAT scores.

It is possible to look at the results of students with similar SAT scores. Within these groups, the letter-grade differences between students with minimal or no high school calculus and those with a year of high school calculus once again emerge.

**Recommendations.** Whatever the recommendations made to high schools about course offerings, the reality is that many students take calculus in high school. We have to address the consequences of this fact. For most of these students, their high school calculus is not sufficient to exempt them from a standard first-year university course, so that they find themselves beginning their college careers by repeating calculus. Their numbers can reach a critical fraction of the population of the introductory college calculus class. When this occurs, not only are the repeating students unchallenged, but students with a standard preparation in high school mathematics can be disadvantaged as the temper of the course adapts itself to an emerging dominant population. The standard introductory course in calculus can then become unprofitable for the one group, and unwholesome for the other.

Calculus, of course, need not be the only entry point to mathematics for the first-year student. Discrete mathematics and elementary statistics and linear algebra can all be suitable courses for the freshman or sophomore student—and genuinely introductory in the sense of leading the student to further study in mathematics. Other worthwhile options for the beginner, such as finite mathematics or the history of mathematics, are often seen as terminal courses. At least one introductory, nonterminal mathematics course should be available to beginning students as an alternative to the calculus.

It is not enough simply to provide alternatives to calculus for first-year students. Students with a sound precalculus preparation in high school should be able to take calculus in college without beginning at a competitive disadvantage. It is not clear at all that these students are unable to learn calculus in the university. What is clear is that they do not compete well, in the same class and at the same speed, with students who began the subject a year ahead of them.

This probably means that we shall have to provide separate beginning calculus courses for students who enter college from the different "calculus tracks" in the high schools. The course for the students with calculus experience could then be adapted so that (as the CUPM panel recommends) it is clearly different from a high school calculus course.

The course for students with a standard precalculus high school preparation should be expected to take a semester longer than its counterpart for students with high school calculus experience. The extra time is important: in any event, it is already being spent by many students simply in repeating their college calculus courses. The two sequences should, at their conclusions, bring students to the same level. We have to live with the effects of a calculus tracking system in the high schools: we should see that the tracks converge somewhere, and the sooner the better.

REFERENCES

1. Ronald G. Douglas, Today's calculus courses are too watered down and outdated to capture the interest of students, *Chronicle of Higher Education*, 34 (Jan. 20, 1988) B1.
2. Report of the CUPM Panel on Calculus Articulation: Problems in the transition from high school calculus to college calculus, this MONTHLY, 94 (1987) 776-785.

# Simple Inequalities And Old Limits

CHUNG-LIE WANG
*Department of Mathematics and Statistics, University of Regina, Regina, Saskatchewan, Canada S4S 0A2*

Let us begin with a couple of definitions. As a common practice in elementary calculus courses (see, for example [**3**, p. 399]), the definition of the logarithmic function $\ln y$ is

$$\ln y = \int_1^y t^{-1}\,dt, \qquad y > 0, \tag{1}$$

The number $e$ is then defined to be the number for which $\ln e = 1$. The exponential function $e^x$ is defined to be the inverse function of $\ln y$.

This paper presents a proof of the limits

$$\lim_{n\to\infty} E_n(x) = e^x \tag{2}$$

and

$$\lim_{n\to\infty} L_n(y) = \ln y, \qquad y > 0, \tag{3}$$

where

$$E_n(x) = \left(1 + \frac{x}{n}\right)^n, \qquad L_n(y) = n\left(y^{1/n} - 1\right), \qquad n \geqslant 1,$$

in a simple and unified manner.

Recently, two interesting articles concerning the limit

$$e = \lim_{n\to\infty} E_n(1) \tag{4}$$

appeared in this MONTHLY [**1**, **2**]. More precisely, C. W. Barnes [**1**] and T. N. T. Goodman [**2**] proved the limit (4) by using the following inequalities

$$e\left(1 + \frac{1}{n}\right)^{-1} \leqslant E_n(1) \leqslant e, \tag{5}$$

which were established through the mean value theorem for integrals with integration by parts and the maximality of the function $t\ln(ne/t)$, $t > 0$, respectively.

We can prove the more general limit (2) by extending (5) to the following:

$$e^x\left(1 + \frac{x}{n}\right)^{-x} \leqslant E_n(x) \leqslant e^x, \quad \text{any real } x. \tag{6}$$

It may be interesting to point out that inequalities similar to (6), for example,

$$e^x\left(1+\frac{x}{n}\right)^{x/2} \leqslant E_n(x) \leqslant e^x, \qquad x>0,$$

were adopted for a similar purpose by Mitrinović [**5**, p. 267].

In order to prove (6), we could simply duplicate Barnes's approach by replacing $(n+1)^{-1}$, $n^{-1}$ with $n(n+x)^{-1}$, $nx^{-1}$, respectively, as the integral limits. Or in Goodman's approach we could introduce a modified function $t\ln(ne/xt)$, $xt>0$. However, the approach to be used here is simpler and different. Before spelling out the details, we note that the sequence $\{L_n\}$ was first studied by G. Hurwitz [**4**]; later, G. H. Hardy [**3**, pp. 143–145, 410–411] gave a rather involved proof of (3) for $y>1$. The case $0<y<1$ was left as an exercise for the reader.

We now proceed with the proof of the limits (2) and (3) by establishing (6) and the inequalities

$$\ln y \leqslant L_n(y) \leqslant y^{1/n}\ln y, \qquad y>0, \tag{7}$$

together with the simple limit

$$\lim_{n\to\infty} y^{1/n} = 1, \qquad y>0. \tag{8}$$

To make the paper self-contained we first note that (8) follows from

$$\begin{aligned} 1 \leqslant y^{1/n} &= \left(1+n\frac{y-1}{n}\right)^{1/n} \\ &\leqslant \left[\left(1+\frac{y-1}{n}\right)^n\right]^{1/n} = \frac{n-1}{n}+\frac{y}{n}, \qquad y\geqslant 1. \end{aligned}$$

In the case $y \leqslant 1$ we use $y^{-1}\geqslant 1$ to complete the argument.

Integrating $t^{-2}\leqslant t^{-1}\leqslant 1$ $(t\geqslant 1)$ from 1 to $z$ in conjunction with (1) we obtain at once

$$1-z^{-1} \leqslant \ln z \leqslant z-1 \tag{9}$$

for $z\geqslant 1$. Since (9) is invariant upon replacing $z$ with $z^{-1}$, (9) is true for all $z>0$.

Finally, setting $z=1+(x/n)$ in (9), a simple manipulation yields

$$e^x \leqslant E_n(x)\left(1+\frac{x}{n}\right)^x \quad \text{and} \quad E_n(x)\leqslant e^x$$

so that (6) follows, while setting $z=y^{1/n}$ in (9), a straightforward transposition yields (7).

It may be interesting and worthwhile to point out that the monotonicity of $E_n(x)$ and $L_n(x)$ in $n$ for any real $x$ are not used in the proofs.

**Acknowledgement.** The author wishes to thank the referee for suggesting the use of A. Hurwitz as a reference.

REFERENCES

1. C. W. Barnes, Euler's constant and $e$, this MONTHLY, 91 (1984) 428–430.
2. T. N. T. Goodman, Maximum products and $\lim(1+1/n)^n = e$, this MONTHLY, 93 (1986) 638–640.
3. G. H. Hardy, A Course of Pure Mathematics, 10th ed., Cambridge University Press, London, 1952.
4. A. Hurwitz, Uber die Einführung der elementaren transzendenten Funktionen in der algebraischen Analysis, *Mathematische Annalen*, 70 (1911) 33–47.
5. D. S. Mitrinović, Analytic Inequalities, Springer-Verlag, Berlin, 1970.

# PROBLEMS AND SOLUTIONS



*For instructions about submitting* ***proposed*** *problems for publication in this department see the inside front cover. Please include solutions, relevant references, etc. Three copies are requested.*

*An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.*

***Solutions*** *should be sent to the address given on the inside front cover. Two copies suffice.*

*A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a* MONTHLY *problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

*For instructions about submitting solutions of Problems, which should be mailed before September 30, 1989, see the inside front cover. Please place the solver's name and mailing address on each (doubled-spaced) sheet. Include a self-addressed card or label if an acknowledgement is desired.*

## ELEMENTARY PROBLEMS

E 3319. *Proposed by Paul Erdös, Hungarian Academy of Science, and János Surányi, Eötvös Loránd University, Budapest.*

Suppose $S$ is a set of $2n + 1$ irrational real numbers. Prove that $S$ has a subset $T$ of cardinality $n + 1$ such that no nonempty subset of $T$ has a rational sum.

E 3320. *Proposed by M. S. Klamkin, University of Alberta, Edmonton.*

Determine positive constants $a$ and $b$ such that the inequality

$$yz + zx + xy \geqslant a\left(y^2z^2 + z^2x^2 + x^2y^2\right) + bxyz$$

holds for all nonnegative $x, y, z$ with $x + y + z = 1$ and is the best possible inequality of this form (in the sense that the inequality need not hold if $a$ or $b$ is increased).

E 3321. *Proposed by Robert Spira, Ashland, OR.*

A *Steiner tree* for a finite set $S$ of points in Euclidean $n$-space is a collection $T$ of straight-line segments such that any two points of $S$ can be joined by a unique path in $T$; its length is the sum of the segment lengths. It is known that the minimum length of a Steiner tree for the corners of a square of side-length one is $1 + \sqrt{3}$. Show that for any positive integer $n$ there exists a Steiner tree of length $1 + (2^{n-1} - 1)\sqrt{3}$ for the set of vertices of an $n$-cube of side-length one.

E 3322. *Proposed by Alexandru Lupaş, Sibiu, Romania.*

Suppose $g$ is an even real-valued function on $[-a, a]$, suppose $g$ is nondecreasing on $[0, a]$, and suppose $h$ is a convex real-valued function on $[-a, a]$. Prove that

$$2a\int_{-a}^{a} g(t)h(t)\,dt \geqslant \int_{-a}^{a} g(t)\,dt\int_{-a}^{a} h(t)\,dt.$$

E 3323. *Proposed by Thane Plambeck, Stanford University, CA.*

A change-making machine gives change in chips of three integer denominations, 1 cent, $\alpha$ cents, and $\beta$ cents, where $1 < \alpha < \beta$. The machine uses the following greedy algorithm: In making $M$ cents change it dispenses chips of value $\beta$ as long as possible, then chips of value $\alpha$ as long as possible, then chips of value 1 to complete the total of $M$. We say that a pair $\alpha < \beta$ is *frugal* if, for every $M$, the machine's algorithm dispenses the minimum number of chips among all ways of making change for $M$ cents with denominations $1, \alpha, \beta$. (For example, $(2, 5)$ is frugal but $(4, 9)$ is not, because the $(4, 9)$ machine dispenses four chips in change for 12 cents instead of the minimum of three.)

For fixed $n > 2$, suppose the values $\alpha < \beta$ are two distinct integers chosen at random from $\{2, 3, \ldots, n\}$ with all pairs equally likely. Obtain an asymptotic formula (when $n$ is large) for the probability that $(\alpha, \beta)$ is frugal.

E 3324. *Proposed by Mark D. Meyerson and R. Bruce Richter, U.S. Naval Academy, Annapolis, MD.*

Suppose that $a$ and $b$ are two closed arcs (homeomorphs of $[0,1]$) in the plane, each of which contains the origin $O$ but not as an endpoint. Prove or disprove: (a) there are arbitrarily small neighborhoods $N$ of $O$ such that $N \cap a$ is connected; (b) there are arbitrarily small neighborhoods $N$ of $O$ such that both $N \cap a$ and $N \cap b$ are connected.

## SOLUTIONS OF ELEMENTARY PROBLEMS

### Some Inequalities for Polynomials

E 2986* [1983, 133]. *Proposed by Abdul Aziz, University of Kashmir, India.*

If $P(z)$ is a polynomial of degree $n$ with complex coefficients having all its zeros in $|z| \geqslant K$, where $K > 1$, prove or disprove the following two assertions:

(i) $$|P(K^2z) - P(z)| \leqslant (K^n - 1)\max_{|z|=1}|P(z)|, \text{ for } |z| = 1.$$

(ii) $$|P(K^2z)| - |P(z)| \leqslant (K^n - 1)|z|^n\max_{|z|=1}|P(z)|, \text{ for } |z| \geqslant 1.$$

*Solution of* (i) *by Paul Ilacqua, Santa Clara, CA.* Assertion (i) is not true in general. To see this take

$$P(z) = (z + K)^p(z - K),$$

where $p$ is a positive integer greater than 1 and $K$ is a real number greater than 1 such that

$$(K^2+1)(p-1) \geqslant 2K(p+1). \tag{1}$$

To determine the maximum of $|P(z)|$ for $z$ on the unit circle it suffices to find the maximum of the following trigonometric polynomial:

$$\begin{aligned} f(\theta) &= |P(e^{i\theta})|^2 = \{(K+\cos\theta)^2+(\sin\theta)^2\}^p\{(K-\cos\theta)^2+(\sin\theta)^2\} \\ &= \{K^2+1+2K\cos\theta\}^p\{K^2+1-2K\cos\theta\}. \end{aligned}$$

Now

$$\begin{aligned} f'(\theta) = &-2K\sin\theta\{K^2+1+2K\cos\theta\}^{p-1} \\ &\times\{(K^2+1)(p-1)-2K(p+1)\cos\theta\}. \end{aligned}$$

On the one hand, $K^2+1+2K\cos\theta \geqslant K^2+1-2K=(K-1)^2>0$ for all real $\theta$ and on the other hand, condition (1) guarantees that $(K^2+1)(p-1)-2K(p+1)\cos\theta$ is positive if $\theta$ is not a multiple of $2\pi$. Hence $f'(\theta)>0$ in $(-\pi,0)$ and $f'(\theta)<0$ in $(0,\pi)$, so that the maximum of $f(\theta)$ occurs when $\theta=0$. Thus the right-hand side of the proposed inequality (i) becomes

$$(K^{p+1}-1)\operatorname*{Max}_{|z|=1}|P(z)| = (K^{p+1}-1)|P(1)| = (K^{p+1}-1)(K+1)^p(K-1).$$

However, when $z=1$ the left-hand side of the proposed inequality (i) becomes

$$\begin{aligned} P(K^2)-P(1) &= (K^2+K)^p(K^2-K)+(K+1)^p(K-1) \\ &= (K^{p+1}+1)(K+1)^p(K-1). \end{aligned}$$

Thus assertion (i) fails for the above choice of $P$ when (1) holds, i.e., when $K \geqslant (\sqrt{p}+1)/(\sqrt{p}-1)$.

*Editorial Comment on* (ii). In a letter to the editors, Professor W. H. J. Fuchs of Cornell University has made the following remarks.

It seems likely that (ii) is true, but we do not have a proof. Lending support to (ii) is the fact that (ii) is correct for $|z|=1$ and, further, that we get a correct assertion if we divide both sides of (ii) by $|z|^n$ and let $|z|\to\infty$.

When $|z|=1$ the inequality (ii) becomes

$$|P(z)|\left\{\prod_{j=1}^{n}\left|\frac{K^2z-A_j}{z-A_j}\right|-1\right\} \leqslant (K^n-1)\max_{|z|=1}|P(z)| \qquad (|z|=1) \tag{2}$$

where $A_1, A_2, \ldots, A_n$ are the zeros of $P$. But if $\rho=z/A_j$, then

$$\left|\frac{K^2z-A_j}{z-A_j}\right| = \left|\frac{1-K^2\rho}{1-\rho}\right| \leqslant K,$$

since $|\rho|\leqslant 1/K$ and

$$K^2-\frac{(1-K^2\rho)(1-K^2\bar{\rho})}{(1-\rho)(1-\bar{\rho})} = \frac{(K^2-1)(1-K^2\rho\bar{\rho})}{(1-\rho)(1-\bar{\rho})} \geqslant 0.$$

Thus the left-hand side of (2) does not exceed $|P(z)|\ (K^n - 1)$ and so (2) is established.

If we divide both sides of (ii) by $|z|^n$ and let $|z| \to \infty$, we get

$$\max_{|z|=1} |P(z)| \geqslant (K^n + 1)|C_n|, \tag{3}$$

where

$$P(z) = C_0 + C_1 z + \cdots + C_n z^n.$$

But (3) follows from the known result

$$\max_{|z|=1} |P(z)| \geqslant |C_0| + |C_n|, \tag{4}$$

which can be proved as follows. Suppose $C_0/C_n = e^{2\pi i\alpha}|C_0/C_n|$. Then

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{n} P\left(e^{2\pi i(j+\alpha)/n}\right) &= \frac{1}{n}\sum_{k=0}^{n} C_k e^{2\pi i\alpha k/n} \sum_{j=1}^{n} e^{2\pi ijk/n} \\ &= e^{2\pi i\alpha} C_n + C_0 \\ &= e^{2\pi i\alpha}(|C_n| + |C_0|)C_n/|C_n|.\end{aligned}$$

Thus (4) is established and hence (3).

A proof of (ii) when $1 < |z| < \infty$ is earnestly solicited.

The above partial solutions were the only ones received.

### Noncrossing Trees

E 3170 [1986, 650]. *Proposed by The Howard University Group, Washington, D.C.*

Construct a graph as follows: Put $n + 1$ labeled vertices around a circle and let the edges be the straight line segments connecting any two vertices. A tree is noncrossing if no two edges intersect except at the vertices. Enumerate the number of noncrossing spanning trees for this graph. For $n = 1, 2, 3$, the numbers are 1, 3, 12, respectively.

*Solution I (composite) by O. P. Lossers, Eindhoven University of Technology, Eindhoven, The Netherlands, and R. S. Pinkham, Stevens Institute of Technology, Hoboken, N.J.* The answer is $\binom{3n}{n-1}/n$. Let $a_n$ denote the required number of trees, and define $a_0 = 1$. We show that

$$a_n = \sum_{j+k+l=n-1} a_j a_k a_l. \tag{1}$$

Label the $n + 1$ vertices $v_0, \ldots, v_n$ in order. Given a noncrossing spanning tree, let $j + 1$ be the smallest label of a vertex connected to $v_0$. In addition to the edge $v_0 v_{j+1}$, the tree consists of a noncrossing tree on $\{v_1, \ldots, v_{j+1}\}$, a noncrossing tree on $\{v_{j+1}, \ldots, v_{j+k+1}\}$ for some $k$ with $0 \leqslant k \leqslant n - j - 1$, and a noncrossing tree on $\{v_{j+k+2}, \ldots, v_n, v_0\}$. Since the noncrossing trees on these vertex sets can be chosen arbitrarily, we have

$$a_n = \sum_{j=0}^{n-1} a_j \sum_{k=0}^{n-j-1} a_k a_{n-j-k-1} = \sum_{j+k+l=n-1} a_j a_k a_l.$$

In general, of course, the number of ballot sequences such that candidate A always has at least $k-1$ times as many votes as candidate B is $\binom{kn+1}{n}/(kn+1)$, if the final tally is $((k-1)n, n)$.

Like the Catalan numbers, these have various interpretations, and there exist bijections between them. In particular, another collection of size $a_n$ that can be used to get bijections between the noncrossing trees and the "2-to-1" ballot sequences is the partitions of $3n$ points on a circle into non-overlapping triangles. The Catalan numbers count the partitions of $2n$ points into noncrossing chords.

Also solved by the proposers. Partially solved by J. T. Ward and the University of South Alabama Problem Group.

### Integers Have Small Multiplicative Order Modulo Some Large Primes

E 3216 [1987, 549]. *Proposed by Jon Froemke and Jerrold W. Grossman, Oakland University, Rochester, Michigan.*

Fix an integer $a > 1$. For any prime $p$ not dividing $a$, let $m(p)$ be the multiplicative order of $a$ modulo $p$, i.e., the smallest positive integer $m(p)$ such that $a^{m(p)} - 1$ is divisible by $p$. Prove that the ratio $(p-1)/m(p)$ is an unbounded function of $p$.

*Solution by O. P. Lossers, Eindhoven University of Technology, Eindhoven, The Netherlands.* Suppose there is an integer $d$ such that $(p-1)/m(p) \leqslant d$ for all primes $p$. By Dirichlet's theorem on primes in arithmetic progressions, we may choose a prime $r$ such that $r > a-1$ and $r \equiv 1 \bmod (d+1)!$. Now there exists a prime $p$ such that $p$ divides $a^r - 1$ but not $a-1$; in fact, any prime dividing $1 + a + a^2 + \cdots + a^{r-1}$ has this property. Since $r$ is prime and $m(p) \neq 1$, we must have $m(p) = r$. By assumption this means $(p-1)/r \leqslant d$, so we conclude $p = kr + 1$ for some integer $k$ with $1 \leqslant k \leqslant d$. However, our choice of $r$ implies $p = kr + 1 \equiv (k+1)(\bmod (d+1)!)$, contradicting the fact that $p$ is prime.

*Editorial comment.* The problem can also be solved quickly via the Tschebotareff density theorem (*Math. Ann.* 95(1926), 191–228). There are other short solutions that are not completely elementary.

Also solved by I. M. Isaacs, R. P. Lewis, L. E. Mattics, S. V. Ullom, and the proposers. See also the proposers' paper in this *Monthly* 95 (1988) 289–307.

### A Property of a Few Small Primes

E 3233 [1987, 877]. *Proposed by Mihai Cipu, INCREST, Bucharest, and Marian Deaconescu, University of Timisoara, Romania.*

Find all primes $p$ having the property that for each prime $q < p$ the least positive residue of $p$ modulo $q$, i.e., $p - \lfloor p/q \rfloor q$, is square free.

*Solution by Lenny Jones, Shippensburg University, PA.* Let $S$ be the set of all primes with the desired property. We will show that $S = \{2, 3, 5, 7, 13\}$. It is easily checked that these primes are in $S$. So consider $p \in S$ with $p > 7$. Then $p - 4$ can have no prime factor $q$ larger than 4, otherwise $p - \lfloor p/q \rfloor q = 4$. Since $p - 4$ is

odd, $p - 4 = 3^a$ with $a \geqslant 2$. Similarly $p - 8$ can have no prime factors larger than 8, and so $p - 8 = 3^a - 4 = 5^b 7^c$. Reduction of the last equation mod 24 implies that $a$ is even and $b$ is odd. If $c \neq 0$, then $p - 9 = 5^b 7^c - 1 = 2^d$ (since $p - 9$ has no prime factor exceeding 8 and is not divisible by $3, 5, 7$); reduction mod 7 implies that this is impossible. Hence $c = 0$ and $3^a - 4 = 5^b$ which, since $3^{a/2} - 2$ and $3^{a/2} + 2$ are relatively prime, gives $3^{a/2} - 2 = 1$ and $3^{a/2} + 2 = 5^b$. Thus $a = 2$, $b = 1$, and $p = 13$.

*Editorial comments* D. Estes and K. S. McCurley provided a short solution based on the following result of K. Molsen (Zur Verallgemeinerung des Bertrandschen Postulates, *Deutsche Math.*, 6 (1941) 248–256): for $n \geqslant 118$, each residue class modulo 12 and relatively prime to 12 contains a prime in the interval $(n, 4n/3]$. Several readers pointed out that $\{2, 3, 5, 7, 13\}$ is the set of primes $p$ having the property that, for each prime $q < p$, the least positive residue of $p$ modulo $q$ is neither 4 nor 9. A few others noted that $\{1, 3, 5, 7, 13\}$ is the set of odd positive integers with this same property. G. Myerson went even further by identifying the entire set of positive integers $p$ that satisfy the property in the statement of the problem. The answer is $\{1, 2, 3, 4, 5, 6, 7, 8, 10, 12, 13, 16, 36, 58\}$. Myerson relies on the paper by L. Alex, Diophantine equations related to finite groups, *Comm. Algebra* 4 (1976) 77–100, in which are given all solutions to $x + y = z$ with $xyz = 2^t 3^u 5^v 7^w$; however, it is possible to give a proof of Myerson's assertion without assuming a knowledge of Alex's paper.

Also solved by D. Callan, D. Estes and K. S. McCurley, S. M. Gagola, C. V. Heuer, H. M. Marston, L. E. Mattics, G. Myerson, S. Philipp, H. Schmidt, D. B. Tyler, and the proposer. Two incorrect solutions were received.

### Too Many Marriages

E 3234 [1987, 877]. *Proposed by Charles R. Johnson, Gail Wolkowicz, and Henry Wolkowicz, University of Delaware, Newark.*

The jury panel for a criminal trial contains 156 members and is chosen from the 3000 registered voters in the small county where the crime occurred. The 3000 voters include 500 married couples, while the jury panel contains $M$ married couples. The defense lawyer argues for a change of venue on the ground that the jury panel was not randomly chosen from the set of registered voters, asserting that the probability of $M$ or more married couples on the jury panel is less than 1 in 20,000. How large must $M$ be for his argument to be correct?

*Solution I by Patrick A. Staley, Southwestern College, Chula Vista, CA.* $M$ must be eight or more. The probability of eight or more couples is less than 1 in 21327, and the probability of seven or more couples is more than 1 in 2986. These numbers were obtained by the following computation.

Given that panel members are selected at random (without replacement), let $P(i, j, k)$ be the probability of a panel consisting of $i$ married couples, $j$ married but unmatched persons and $k$ single persons after $2i + j + k$ picks. Set $P(i, j, k) = 0$ if any argument is negative, and set $P(0, 0, 0) = 1$. Then $P(i, j, k)$ can be

computed using the recurrence

$$\begin{aligned}(3001 - 2i - j - k)P(i, j, k) = {} & (j+1)P(i-1, j+1, k) \\ & + (1002 - 2j - 2i)P(i, j-1, k) \\ & + (2001 - k)P(i, j, k-1)\end{aligned}$$

The resulting probability of $M$ or more couples is then

$$\sum_{i=M}^{78} \sum_{j=0}^{156-2i} P(i, j, 156 - 2i - j).$$

*Solution II by Curtis Cooper, Central Missouri State University.* $M$ must be eight or more. Assuming random selection, the probability that the jury panel has exactly $i$ married couples and exactly $j$ people who are married but whose spouse is not chosen is

$$P(i, j) = \frac{\binom{500}{i}\binom{500-i}{j}2^j\binom{2000}{156-2i-j}}{\binom{3000}{156}}.$$

As above, we compute $\sum_{i=M}^{78}\sum_{j=0}^{156-2i}P(i, j)$ and compare with $1/20000$.

*Editorial comment.* Lajos Takacs remarked that by the principle of inclusion-exclusion, the probability $p$ of $M$ or more married couples satisfies

$$p = \sum_{j=M}^{78} (-1)^{j-M}\binom{j-1}{M-1}\binom{500}{j}\frac{\binom{156}{2j}}{\binom{3000}{2j}},$$

which can be readily calculated for $M = 7, 8$. (cf. Tucker, *Applied Combinatorics*, New York, Wiley, 1984, p. 313).

Also solved by D. Callan, J. Delany, P. Griffin, and the proposers. One incorrect solution was received.

### A Sensible Allocation of Funds?

E 3235 [1987, 877]. *Proposed by Rick Luttmann, Sonoma State University, Rohnert Park, CA.*

In the primary of June 1986, the voters of the state of California were asked to vote on a bond issue that contained the following language: "The total funds available shall be allocated to counties.... Each county's allocation shall be in the same ratio as the county's population is to the state's total population, except that each county shall be entitled to a minimum allocation of $100,000."

While this language was apparently not written by mathematicians, it presumably means the following:

(a) Each allocation shall be at least $100,000.

(b) Those counties that receive more than $100,000 shall have equal per capita allocations.

(c) The per capita allocation of counties that receive only $100,000 shall be not less than the common per capita allocation of those counties in (b).

Let $T$ be the total of funds available, $m$ the minimum allocation, and $c$ the number of counties. Let $p_1, p_2, \ldots, p_c$ denote the counties' populations, arranged in descending order. Find a formula for each county's allocation.

*Solution by Robert A. Agnew, Chicago.* Let $x_i$ be the allocation of the $i$th county, $n$ the number of counties receiving more than the minimum, and $a$ their common per capita allocation. Then $x_i = m$ for $i > n$ (if $n < c$), and $x_i = ap_i \geqslant m$ for $i \leqslant n$ (by (b) and (c)). Since $\Sigma x_i = T$, we obtain

$$a = \frac{T - (c-n)m}{\sum_{i=1}^{n} p_i}.$$

Furthermore,

$$n = \max\left\{ k \leqslant c \colon \frac{T-(c-k)m}{\sum_{i=1}^{k} p_i} \geqslant \frac{m}{p_k} \right\}.$$

If $cm > T$, then there is no solution.

Note that for any constant $b$, the solution to the proposed problem also solves the simple weighted least-squares problem of minimizing $\sum_{i=1}^{c} p_i (x_i/p_i - b)^2$ subject to $\sum_{i=1}^{c} x_i = T$ and $x_i \geqslant m$ for all $i$. This can be established via the Kuhn-Tucker conditions.

Also solved by S. F. Barger, J. R. Buchanan, D. Callan, J. Gaisser, A. Gorfin, C. Groenewoud, E. Hertz, P. Montuschi (Italy), W. H. Oh (Korea), A. Riese, V. Schindler (East Germany), M. Shimshoni (Israel), A. Shuchat, and the proposer.

**A Bernoulli Recurrence**

E 3237 [1987, 995]. *Proposed by Johan G. F. Belinfante, Georgia Institute of Technology, Atlanta.*

Let $B_k$ denote the $k$th Bernoulli number, as defined by the formal power series expansion

$$\frac{x}{e^x - 1} = \sum_{k=0}^{\infty} B_k \frac{x^k}{k!}.$$

Prove the following generalization of a formula of Namias [*Amer. Math. Monthly*, 93 (1986) 28–29]:

$$B_m = \frac{1}{n(1-n^m)} \sum_{k=0}^{m-1} n^k \binom{m}{k} B_k \sum_{j=1}^{n-1} j^{m-k},$$

valid for any positive integer $m$ and any positive integer $n$ greater than 1.

*Solution by Ira Gessel, Brandeis University, Waltham, MA.* Let $B(x) = x/(e^x - 1) = \Sigma B_k x^k/k!$. Then

$$\frac{B(nx)}{n} \sum_{j=0}^{n-1} e^{jx} = \frac{x}{e^{nx}-1} \frac{e^{nx}-1}{e^x - 1} = B(x).$$

Equating coefficients of $x^m/m!$ after expressing the left side as a product of two generating functions yields the formula

$$\frac{1}{n}\sum_{k=0}^{m} n^k\binom{m}{k}B_k\sum_{j=0}^{n-1} j^{m-k} = B_m,$$

where $0^0 = 1$. Subtracting the term $k = m$ and normalizing yields the desired formula.

Also solved by E. Y. Deeba and D. M. Rodriguez, O. P. Lossers (The Netherlands), S. Philipp, O. G. Ruehr, the University of South Alabama Problem Group, and the proposer.

**An Entirely Real Function**

E 3242 [1987, 996]. *Proposed by J. B. Miles and L. A. Rubel, University of Illinois, Urbana.*

It can be proved, using Picard's theorem or Nevalinna theory, that if $f(z)$ is an entire function that is real precisely when $z$ is real, then $f(z) = az + b$, where $a$ and $b$ are real numbers and $a \neq 0$. Find a more elementary proof.

*Solution I by Stan Philipp, Pennsylvania State University, Altoona.* Let $f(z) = \sum a_n z^n = u(z) + iv(z)$, with $u$ and $v$ real. Note that the $a_n$ are real, since $f$ is real for real $z$. Since $v(z) \neq 0$ for $\operatorname{Im}(z) \neq 0$, $v$ must have constant sign in the upper half plane. Therefore, using the inequality $|\sin n\theta| \leqslant n\sin\theta$ for $n = 1, 2, \ldots$ and $0 \leqslant \theta \leqslant \pi$ and the formula

$$\int_0^\pi v(re^{i\theta})\sin m\theta\, d\theta = \sum_{n=1}^{\infty} a_n r^n \int_0^\pi \sin n\theta \sin m\theta\, d\theta = \frac{\pi}{2}a_m r^m \quad \text{for } m \geqslant 1,$$

we find

$$|a_m| \leqslant \frac{2}{\pi r^m}\int_0^\pi \left|v(re^{i\theta})\right| m\sin\theta\, d\theta = \frac{m|a_1|}{r^{m-1}}$$

for any positive $r$ and $m \geqslant 1$. Letting $r \to \infty$ shows $a_m = 0$ for $m \geqslant 2$. Thus $f(z) = a_0 + a_1 z$. Since $f$ is not real off the real axis, $a_1 \neq 0$.

*Solution II by O. P. Lossers, Eindhoven Institute of Technology, Eindhoven, The Netherlands.* We may assume without loss of generality that $\operatorname{Im} f > 0$ on the upper half plane and then, since $f(\bar{z}) = \overline{f(z)}$, that $\operatorname{Im} f < 0$ on the lower half plane.

Also, $f'$ has no zeros on the real axis. Otherwise, if $f'(x) = 0$ for some $x \in \mathbb{R}$, we would have

$$f(z) = f(x) + f^{(k)}(x)\frac{(z-x)^k}{k!}(1 + h(z))$$

for some $k \geqslant 2$, where $f^{(k)}(x) \in \mathbb{R}$ and $h$ is an entire function with $h(x) = 0$. This implies that in any neighborhood of $x$ there is some $z$ with $\operatorname{Im} z > 0$, such that $f^{(k)}(x)(z-x)^k/k!$ is on the negative imaginary axis, and $z$ is close enough to $x$ so that $\operatorname{Re}(1 + h(z)) > 0$. Hence $\operatorname{Im} f(z)$ is negative, contradicting our assumption.

We conclude that $f$ is strictly increasing on the real axis. Now consider the function $g(z) = (f(z) - f(0))/z$. This function is entire and satisfies $g(x) > 0$ for

$x \in \mathbb{R}$ and $|\arg(g(z))| < \pi$ for all $z$, since $\operatorname{sgn}(\operatorname{Im} f(z)) = \operatorname{sgn}(\operatorname{Im} z)$. Therefore, the function $G(z) = 1/(1 + \sqrt{g(z)})$ is a bounded entire function and hence constant. Hence $g$ is constant, which proves the assertion.

*Editorial Comment.* Dale H. King mentions related results of E. P. Wigner, "Simplified Derivation of Properties of Elementary Transcendentals," *Amer. Math. Monthly*, 59(1952) 669–683. H. P. Boas remarks that the result of the problem is an immediate corollary of the following proposition: *A positive harmonic function $v$ in the upper half-plane that is zero on the real axis must have the form $v = ay$.* This proposition occurs both as Theorem 8.34 in Robert B. Burckel, *An Introduction to Classical Complex Analysis I* (Academic Press, 1979) and Theorem 1 in H. P. Boas and R. P. Boas, "Short Proofs of Three Theorems on Harmonic Functions," *Proc. Amer. Math. Soc.*, 102(1988) 906–908. The result of the problem goes back at least to Tschebotareff, *Math. Ann.* 99 (1928) 675–677.

Solved also by N. K. Artemiadis (Greece), M. R. Aub (Jamaica), H. P. Boas, F. W. Carroll, R. B. Israel (Canada), D. H. King, K.-W. Lau (Hong Kong), E. Levine, L. E. Mattics, P. Tracy, and the proposers.

## ADVANCED PROBLEMS

6598. *Proposed by Walter Rudin, University of Wisconsin, Madison.*

Let $K$ be a closed circular disc in the plane $R^2$. Suppose that $f$ is a continuous map from $K$ into $R^2$ which is one-to-one on the interior of $K$. For each point $q$ in $f(K)$, let $N(q)$ denote the cardinality of the set of all points $p$ in $K$ for which $f(p) = q$.

Prove that $N(q) \geqslant 3$ for at most countably many points $q$.

6599. *Proposed by Nick MacKinnon, Winchester College, Winchester, England.*

Let $m(n)$ denote the sum of the prime factors of $n$, repetitions counting. For example $m(12) = 2 + 2 + 3 = 7$. What is the asymptotic behavior of

$$\sum_{n=1}^{N} m(n)$$

for large $N$?

6600. *Proposed by P. J. de Doelder, Technical University, Eindhoven, The Netherlands.*

If $m$ is a positive integer, prove that

$$\int_0^{\pi/2} \frac{u^{2m}}{\sin u}\,du = 2(-1)^m(2m)!\,\lambda(2m+1)$$
$$+2(-1)^{m+1}(2m)!\sum_{j=0}^{m-1}\frac{(-1)^j(\pi/2)^{2j+1}}{(2j+1)!}\beta(2m-2j)$$

and

$$\int_0^{\pi/2} \frac{u^{2m-1}}{\sin u}\, du = 2(-1)^{m+1}(2m-1)! \sum_{j=0}^{m-1} \frac{(-1)^j (\pi/2)^{2j}}{(2j)!} \beta(2m-2j),$$

where

$$\lambda(s) = \sum_{k=0}^{\infty} \frac{1}{(2k+1)^s} \qquad \beta(s) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)^s}.$$

## SOLUTIONS OF ADVANCED PROBLEMS

6531 [1986, 821]. *Proposed by R. H. Jeurissen, Katholieke Universiteit, Nijmegen, The Netherlands.*

Let $G$ be the (bipartite) graph of which the points are the elements of the symmetric group $S_n$, with $\sigma$ and $\tau$ adjacent if and only if there is an $i, 0 < i < n$, with $\sigma = \tau(i, i+1)$, i.e., the sequence $\sigma(1), \sigma(2), \dots, \sigma(n)$ can be transferred into $\tau(1), \tau(2), \dots, \tau(n)$ by interchanging two consecutive elements. Determine the automorphism group of $G$ (cf. problem 6486 [1986, 62; 1986, 574]).

*Solution by the proposer.* Clearly for all $\tau \in S_n$, the transformations $L_\tau$ given by

$$L_\tau\colon \sigma \to \tau\sigma$$

are automorphisms. Thus Aut $G$ contains a transitive subgroup isomorphic to $S_n$, and hence $G$ is regular. If $n > 2$ another automorphism is the involutory conjugation $C$ by $(1, n)(2, n-1) \cdots (\lfloor \frac{1}{2}n \rfloor, \lceil \frac{1}{2}n \rceil + 1)$. We shall prove that Aut $G$ consists of the $L_\tau$ and (if $n > 1$) the $L_\tau \circ C$, so if $n = 1, 2$ we have Aut $G \simeq S_n$, while if $n > 2$, then Aut $G \simeq C_2 \cdot S_n$, a semidirect product ($L_\tau \circ C = C \circ L_{C\tau}$). The cases $n = 1, 2$ are trivial, so we suppose $n > 2$.

Let $T_n = \{(i, i+1) | i = 1, 2, \dots, n-1\}$ be the set of neighbours of (1). (Since $T_n$ generates $S_n$, the graph $G$ is connected.) The neighbours other than 1 of $(i, i+1)$ are the $(i, i+1)(j, j+1)$ with $i+1 < j < n$ or $0 < j < i-1$, $(i, i+1, i+2)$ if $i < n-1$, and $(i-1, i+1, i)$ if $1 < i$. Let $Q_n$ be the set of neighbours of elements of $T_n$ of the first kind, $R_n$ the set of the neighbours of the other two types. The points of $Q_n$ each have 2 neighbours in $T_n$, those of $R_n$ each 1; (1 2) and $(n-1, n)$ each have 1 neighbour in $R_n$, the other elements each have 2. It follows that any automorphism $\phi$ fixing (1), fixes $T_n$, $Q_n$, and $R_n$, and fixes (1 2) and $(n-1, n)$ or interchanges them. Let $\phi' = \phi$ if $\phi$ fixes (1 2), $\phi' = C \circ \phi$ otherwise. Then $\phi'$ fixes (1) and (1 2), hence also (2 3), this being the only element of $T_n$ sharing no neighbour in $Q_n$ with (1 2). Similarly we see that $\phi'$ fixes (3 4), $\dots, (n-1, n)$. We now have: if $\psi$ is an automorphism, and $\psi((1)) = \sigma$, then either $L_{\sigma^{-1}} \circ \psi$ or $C \circ L_{\sigma^{-1}} \circ \psi$ fixes (1) and every element of $T_n$, and it suffices to prove that an automorphism with this property is the identity.

So suppose $\phi$ is an automorphism distinct from the identity that fixes (1) and all $(i, i+1)$. Let $\phi(\sigma) \neq \sigma$, where $\sigma$ is chosen such that its distance $k$ from (1) is minimal. Then $k > 1$, and $\sigma$ is of the form $\tau(j, j+1)$ with $\tau$ at distance $k-1$ from (1) and $\phi(\tau) = \tau$. Let $\psi = L_{\tau^{-1}} \circ \phi \circ L_\tau$: then $\psi((1)) = (1)$. Thus either $\phi$ fixes

$(i, i+1)$ for all $i$ or $\psi((i, i+1)) = C((i, i+1))$. In the first case, however, we would have $\phi(\tau(i, i+1)) = \tau(i, i+1)$ for all $i$, in particular $\phi(\sigma) = \sigma$. So we have $\phi(\tau(i, i+1)) = \tau C((i, i+1) = \tau(n-i, n+1-i)$ for all $i$; therefore a neighbour $\tau(i, i+1)$ of $\tau$ is invariant iff $i = \frac{1}{2}n$. Since indeed $\tau$ must have an invariant neighbour (at distance $k-2$ from (1)), we find that $n$ is even, that $\rho = \tau(\frac{1}{2}n, \frac{1}{2}n+1)$ is the only neighbour of $\tau$ at distance $k-2$ from (1), and that all other neighbours are at distance $k$ from (1) and are not invariant. In particular $\tau' = \tau(1\,2)$ and $\tau'' = \tau(n-1, n)$ are interchanged. For the moment let $n \geqslant 6$. Then $\rho(1\,2) = \tau(\frac{1}{2}n, \frac{1}{2}n+1)(1\,2) = \tau(1\,2)(\frac{1}{2}n, \frac{1}{2}n+1)$ is a neighbour of both $\rho$ and $\tau'$, so it is at distance $k-1$ from (1) and invariant. So it should be a neighbour of $\tau'' = \phi(\tau'')$ too, i.e., there is a $j$ with $(n-1, n)(j, j+1) = (1\,2)(\frac{1}{2}n, \frac{1}{2}n+1)$. This is impossible. So $n = 4$. Now $\tau'$ and $\tau''$ have $\tau$ and also $\tau(1\,2)(3\,4)$ as common neighbours, so their third neighbours $\xi = \tau(1\,2)(2\,3)$ and $\eta = \tau(3\,4)(2\,3)$ are interchanged. The element $\xi(1\,2)(2\,3) = \rho(1\,2)$, however, is at distance $\leqslant k-1$ from (1), so it is invariant, and at distance 2 from $\xi$. So it is at distance 2 from $\eta$. This implies that there are $i$ and $j$ with $(2\,3)(1\,2)(i, i+1)(j, j+1)(3\,4)(2\,3)$, or $(i, i+1)(j, j+1) = (1\,2\,4)$. This is impossible.

*Editorial Comment.* Professor John H. Walter of the University of Illinois has made the following remarks.

The question stated in this problem is a well-known and often used result about reflection groups used in the theory of Chevalley groups. An outline of its connection to this theory may be of interest. These groups are finite Euclidean reflection groups of $\mathbb{R}^n$ and have a presentation $G = \langle \tau_i, i = 1, 2, \ldots, n | (\tau_i\tau_j)^{m_{ij}} = 1, m_{ii} = 2\rangle$. The possibilities for the integers $m_{ij}$ are given by the Dynkin diagram $\Delta(G)$. It is well known that the symmetric group $S_{n+1}$ is associated with diagrams of type $A_n$. The graph considered in this problem is the Cayley graph $\Gamma(G)$. It describes the adjacency structure of the fundamental regions of $G$ acting on $\mathbb{R}^n$, and $\operatorname{Aut}(\Gamma(G))/G$ is the group of symmetries of $\Delta(G)$. When $\Delta(G)$ has type $A_n$, this symmetry group has order 2.

Besides the standard treatments of the theory of Euclidean reflection groups, reference is also made to the more comprehensive treatment of this subject from the point of view of thin chamber systems by Jacques Tits in *Buildings of Spherical Type and Finite BN-pairs*, Lecture Notes in Mathematics, v. 386, Springer, 1974.

Also solved by O. P. Lossers (The Netherlands).

### An Identity of Christoffel

6547 [1987, 469]. *Proposed by P. L. Butzer and the late E. L. Stark, Rheinisch-Westfalische Technischen Hochschule, Aachen, West Germany.*

In discussing the nondifferentiability of the famous "Riemann function" $\sum n^{-2}\cos(n^2x)$, Christoffel in a letter to E. Prym of June 18, 1865 states that

$$\sum_{n=1}^{\infty} \frac{\cos(n^2x)}{n^2} = \frac{\pi^2}{6} - \frac{1}{2}\left(\frac{\pi x}{2}\right)^{1/2} + \frac{\pi^{3/2}}{2^{1/2}} \sum_{n=1}^{\infty} n \int_{n^2\pi^2/x}^{\infty} \frac{\sin t - \cos t}{t^{3/2}}\, dt.$$

Show that, except for the coefficient of the term in $x^{1/2}$, this formula is correct. The text of Christoffel's letter is given in the article by the proposers, "'Riemann's example' of a continuous nondifferentiable function in the light of two letters (1865)

of Christoffel to Prym," Professor Guy Hirsch Birthday Volume, *Bull. Soc. Math. Belg.*, 38A (1986) 45–73.

*Solution by Harold G. Diamond, University of Illinois at Urbana-Champaign.* The formula of the problem is correct except for a superfluous factor $\frac{1}{2}$ in the second term on the right-hand side; that term should be simply $-(\pi x/2)^{1/2}$. This modified assertion is a consequence of the famous theta inversion formula, that

$$\sum_{n=-\infty}^{\infty} e^{-\pi n^2 z} = \frac{1}{\sqrt{z}} \sum_{m=-\infty}^{\infty} e^{-\pi m^2/z}$$

for $\operatorname{Re}(z) > 0$. If we divide by 2 and then integrate the left-hand side from $\varepsilon$ to $\varepsilon + iy$, where $\varepsilon > 0$, we obtain

$$L_\varepsilon(y) = iy/2 + \sum_{n=1}^{\infty} \frac{e^{-n^2\pi(\varepsilon+iy)} - e^{-n^2\pi\varepsilon}}{(-n^2\pi)}.$$

Note that

$$\lim_{\varepsilon\to 0^+} L_\varepsilon(y) = iy/2 + \frac{\pi}{6} - \sum_{n=1}^{\infty} \frac{e^{-in^2\pi y}}{\pi n^2}.$$

The corresponding integral of the right-hand side is

$$\begin{aligned} R_\varepsilon(y) &= \int_\varepsilon^{\varepsilon+iy} \left( \frac{1}{2\sqrt{z}} + \frac{1}{\sqrt{z}} \sum_{n=1}^{\infty} e^{-n^2\pi/z} \right) dz \\ &= \sqrt{\varepsilon + iy} - \sqrt{\varepsilon} + \sum_{n=1}^{\infty} \int_0^y \frac{1}{\sqrt{\varepsilon + it}} e^{-n^2\pi/(\varepsilon+it)} i\, dt. \end{aligned}$$

For fixed $y > 0$ the last series, briefly $\sum_{n=1}^{\infty} a_n(\varepsilon, y)$, converges uniformly for $\varepsilon > 0$. Indeed

$$a_n(\varepsilon, y) = \frac{1}{n^2\pi} \int_0^y (\varepsilon + it)^{3/2}\, d\big(\exp\{-n^2\pi/(\varepsilon + it)\}\big),$$

so that integration by parts and the estimate

$$|\exp\{-n^2\pi/(\varepsilon + it)\}| \leqslant 1$$

imply that $a_n(\varepsilon, y) = O(n^{-2})$ for fixed $y$. The change of variable $t = n^2\pi/u$ shows that

$$a_n(\varepsilon, y) = i\int_{n^2\pi/y}^{\infty} \sqrt{\frac{u}{\varepsilon u + in^2\pi}} \exp\left( \frac{-n^2\pi u}{\varepsilon u + in^2\pi} \right) \frac{n^2\pi}{u^2}\, du.$$

Hence,

$$\lim_{\varepsilon\to 0^+} R_\varepsilon(y) = \sqrt{iy} + \sqrt{\pi} \sum_{n=1}^{\infty} n \int_{n^2\pi/y}^{\infty} \frac{1+i}{\sqrt{2}} e^{iu} u^{-3/2}\, du.$$

Upon equating the real parts of these two limits we obtain

$$\frac{\pi}{6} - \sum_{n=1}^{\infty} \frac{\cos(\pi n^2 y)}{\pi n^2} = \sqrt{(y/2)} + \sqrt{(\pi/2)} \sum_{n=1}^{\infty} n \int_{n^2\pi/y}^{\infty} (\cos u - \sin u) \frac{du}{u^{3/2}};$$

the result follows upon replacing $y$ by $x/\pi$.

No other solutions were received.

# REVIEWS

EDITED BY JOSEPH KONHAUSER
*Department of Mathematics, Macalester College, St. Paul, MN 55105*

*Geometries and Groups.* By V. V. Nikulin and I. R. Shafarevich. Translated from the Russian by M. Reid. Springer-Verlag Universitext, 1987. viii + 251 pp.

HEINRICH W. GUGGENHEIMER
*Department of Mathematics, Polytechnic University of New York, Farmingdale, NY 11735*

From its beginnings, geometry was the meeting ground of topology and algebra. At least, that was the attitude of Euclid. For example, Euclid **[H, T]** is very careful in Definition 3 to assert that a segment contains its endpoints (and not as Heath **[H]** puts it, "if a line *has* extremities, those extremities are points"). This means that all segments (a.k.a. "lines"), just as all other geometric figures appearing in Euclid, are compact. Straight lines in the modern sense, as unbounded point sets, never appear in Euclid. (They do appear in Proclus and Pappus, but these later authors were quite oblivious to the topological insights contained in Euclid.) As a consequence, whenever Euclid needs a point $C$ not on the segment $AB$ which defines the abstract underlying line, he first has to *produce* (i.e., prolong) the segment, i.e., embed it in a larger segment. The unlimited possibility of such prolongation is asserted in Postulate 1. The restriction of geometric figures to compact sets justifies Axiom 8 in the Axioms of Measurement: the whole is greater than its part. This means that if $A$, $B$ are compact and $A$ is strictly contained in $B$, then $m(A) < m(B)$ for the measure $m$ which is invariant under the group of Euclidean motions (Axiom 7). We find the same insistence on compact sets in the Parallel Postulate 5: *And, if a segment intersecting two segments makes interior angles on one side that sum to less than two right angles then the two segments can be produced to intersect on the side that contains the angles that add to less than two right angles.* This postulate contains three statements: Pasch's Axiom on half planes, the Euclidean Axiom on the existence of a unique parallel, and the statement that the point of intersection of intersecting lines can be found in a compact subset of the plane **[G]**. In his careful attention to topological matters, Euclid had no successor and no adequate modern commentator.

The algebra in Euclid comes in two forms. First as geometrical algebra in Books II, V, X and then in the study of the groups of motions and similitudes. The latter naturally was not recognized as algebra until the middle of the nineteenth century.

The story of Euclidean geometry and the discovery of non-Euclidean geometry by Lobachevskii (1829) and Bolyai (1832) has been told many times, so we can continue our story after 1832. Both Lobachevskii and Bolyai worked within the framework of synthetic geometry as taught in the schools. A few years earlier, Poncelet **[P]** had developed projective geometry and naively introduced projective coordinates over the complex numbers. (The use of the complexification of real algebraic curves in projective geometry was justified in 1847 by von Staudt **[vS]**.) Poncelet had noted that in projective coordinates, given by $x = x_1/x_0$, $y = x_2/x_0$, every circle in the plane passes through the two *circular points* $(0, 1, \pm i)$ and had

made some applications of this notion. In 1853, Laguerre **[La]** noted that any theorem about angles can be transformed into a projective theorem if the angle is defined as $(i/2)$ times the logarithm of the cross ratio of the two legs of the angle and the two isotropic lines (of slope $\pm i$) through the vertex. This result was the starting point for A. Cayley (1859) who replaced the degenerate conic $x_1^2 + x_2^2 = 0$ of Euclidean geometry by a nondegenerate conic which then is used as basis of projective definitions both of length and angle **[Ca]**. Depending on the sign of the discriminant of the quadratic form he obtains elliptic or hyperbolic (Lobachevskii-Bolyai) geometry. Elliptic geometry is also called Riemannian geometry because Riemann had in his *Habilitationsvortrag* (lecture delivered at Göttingen University to obtain the right to be an (unpaid) lecturer at that institution) sketched this geometry based on the topology on manifolds and Riemannian geometry, both introduced in a few paragraphs **[R]**. Riemann also indicated that a manifold with constant curvature is a homogeneous space (in modern terminology), i.e., it admits a manifold of isometries in itself and he pointed out that the sphere is a local model of elliptic geometry. Based on Riemann's ideas, Beltrami **[B]** in 1868 found surfaces in three-space whose inner geometry provides a model for local hyperbolic geometry.

Much of modern algebra and geometry goes back to C. Jordan's immersion in the thought of E. Galois and his group theoretical treatment of groups in geometry **[J1]** (1868) and algebra **[J2]** (1870). The study of groups in geometry was immediately taken up by S. Lie (who was in France in 1871) and his friend F. Klein. Lie from 1871 on developed the theory of groups whose underlying space is a manifold and whose multiplication is analytic in local coordinates, known as Lie groups. Klein in 1872 gave a definition of geometry as the study of invariants of homogeneous spaces **[K1]** but concentrated on the study of discrete groups, in particular subgroups of the Möbius group that are important in complex analysis. At the same time, W. K. Clifford in England started to study the groups of motions of non-Euclidean geometry and in 1873 **[C1]** discovered a quadric surface whose geometry in elliptic three-space is locally Euclidean but whose topology in the large is not that of the euclidean plane. The Clifford surface as a whole admits only a two-parameter group of motions whereas the neighborhood of any point admits a three-dimensional group germ.

In the two decades following, S. Lie developed his theory of continuous groups. In modern language these are Lie group germs and the corresponding Lie algebras. Lie tried to classify his groups and pointed out the importance of simple groups in this endeavor. However, he took as his principle of classification the dimension of group space or of the space on which the group may act as a transitive group of transformations and this is not the right way to go. The right way of classification was found by W. Killing, a high school teacher in Braunsberg, who, with remarkable vision in a series of papers from 1887 to 1889 **[Ki1]**, built the modern theory of Lie algebras and Lie groups and gave the complete and almost correct classification (with defective proofs); the details were later put in order by E. Cartan. (In the preface of Vol. 3 of Lie's Transformationsgruppen, Killing is considered to have "deficient knowledge in the theory of groups.")

In 1875, Klein had started to look for all discrete subgroups of the Möbius group and their geometries. One of the first discoveries he made was that there are groups

that leave a circle invariant and others without such a circle **[K2]**. Groups of the first kind were encountered also by Fuchs **[F]**, who in 1880 studied the question of finding those analytic linear equations $y'' + p(x)y' + q(x)y = 0$ with rational $p$, $q$, for which $x$ is a (univalent) function of the quotient $y_2/y_1$ of two fundamental solutions. Poincaré, on reading Fuchs's paper, got interested in the general question of automorphic functions, i.e., analytic functions invariant under a discrete subgroup of the Möbius group. After the publication of his first note on the subject **[Pc1]**, Klein pointed out to him the existence of the second kind of group. Poincaré proceeded to call the first kind *Fuchsian*, the second kind *Kleinean*, and these names have stuck to this day. In his investigations he found the Poincaré model $H$ of hyperbolic geometry and a classification principle for all Fuchsian groups $G$ such that the geometry obtained by identifying points of $H$ that are images of one another under $G$, is locally hyperbolic **[Pc2]**. He noted that the geometry so obtained is uniform (one neighborhood looking like any other, or the underlying space being a manifold) only if $G$ has no fixed point in its action on $H$. In 1890, when Klein turned from original research to the writing of expository books, he noted that Poincaré had given the basic tools for the solution of the problem that followed from Clifford's discovery: To find all geometries that are locally isomorphic to a given geometry defined on a simply connected manifold. He solved the two- and three-dimensional Euclidean and elliptic cases. (The general elliptic case was put in order only in the Ph.D. thesis of H. Hopf **[Ho]**). Killing **[Ki2]** immediately treated the problem from the point of view of Riemannian geometry and Lie groups in arbitrary dimensions and gave a complete solution in the low-dimensional Euclidean and elliptic cases and a reasonably satisfactory one for the still-open hyperbolic case in dimensions two and three. He also coined the name "Clifford-Klein space problem." (Killing had written a few earlier books on "$n$-dimensional non-Euclidean space forms" (Braunsberg 1883, Leipzig 1885) but there "space form" means "Lie group of transformations.") A related problem, the determination of all (Euclidean) crystal classes, had been treated incompletely by Jordan **[J1]** and completely by Fëdorov **[Fe]** in 1890 and independently by Schönfliess **[S]** in 1891.

The modern formulation of the notion of geometry and of the Clifford-Klein space problem has its base in the early work of E. Cartan [C]. Let $G$ be a Lie group and $H$ a closed subgroup. Then the (left) coset space $G/H$ is a *homogeneous space*, i.e., a topological space on which $G$ acts as a transitive group of transformations and the inner isomorphs of $H$ are the isotropy groups, i.e., the groups of transformations that leave a certain point fixed since $gHg^{-1}$ transforms $gH$ into itself. To get a nice quotient space, one has to insist on some additional properties such as "$H$ should not contain a normal subgroup of $G$ of dimension $> 1$." The Clifford-Klein problem can then be formulated as the search for all $K$ such that a simply connected $G/H$ is the universal covering space of $(G/H)/K$. In this way, algebra and topology have finally merged in the notion of Lie group to define geometry.

The book of Nikulin and Shafarevich gives a very careful, complete, and leisurely introduction to the results of Klein and Poincaré on the Clifford-Klein space problem in two and three dimensions and of Jordan-(Fëdorov-Schönfliess) in the plane, with attention to some aspects not covered in this historical essay, such as the problem of moduli of the flat torus which opens up vistas of algebraic geometry. The book is well written and edited (I found only 3 typos). It would serve admirably

for a one-semester geometry course in a liberal arts college since its prerequisite is only a moderate amount of high school geometry. All topology and algebra is developed in the text. Even complex numbers are developed from first geometric principles. For maximum effect, the teacher should obviously know somewhat more than is presented in the text, and here the most important source [M] should be added on top of the list of recommended advanced readings. (The book on Riemannian geometry on the list is obviously by Elie Cartan and not by Henri Cartan as listed.) If you have a full year to spend on an elementary course that can whet someone's appetite for research, take Magnus.

REFERENCES

[B] E. Beltrami, Saggio di interpretatzione della geometria non-euclidea, *Giorn. di Mat.* (Battaglini), 6 (1868) 285–315, French translation by Hoüel, *Ann. Ec. Norm.*, 6 (1868) 251–288.

[C] E. Cartan, Sur la structure des groupes de transformations finis et continus, Ph.D. Thesis, Paris 1984, Oeuvres, vol. I/1, 137–287.

[Ca] A. Cayley, A sixth memoir on quantics, *Lond. Trans.*, 149 (1859) 61–90; Collected Math. Papers vol. 2, 561–592.

[C1] W. K. Clifford, Preliminary sketch of biquaternions, *Proc. London Math. Soc.*, 4 (1873) 381–395; Math. papers 181–200.

[Fe] E. V. Fëdorov, Symmetry of Regular Systems of Figures (in Russian), St. Petersburg, 1890.

[F] L. Fuchs, Sur une classe de fonctions de plusieurs variables tirées de l'inversion des intégrales de solutions des équations différentielles linéaires dont les coefficients sont des fonctions rationelles. Extrait d'une lettre adressée à M. Hermite. C. R. Paris 90 (1880) 678–680, 735–736; Werke II, 213–218.

[G] H. W. Guggenheimer, The axioms of betweenness in Euclid, *Dialectica*, 31 (1977) 187–192. MR 81h:03029.

[H] T. L. Heath, The Thirteen Books of Euclid's Elements, reprint, Dover, N.Y., 1956.

[Ho] H. Hopf, Zum Clifford-Kleinschen Raumproblem, *Math. Ann.*, 95 (1925) 313–339.

[J1] C. Jordan, Mémoire sur les groupes de mouvements, *Ann. di mat.*, (2)2 (1868/69) 167–215, 322–345.

[J2] ______, Traité des substitutions et des équations algébriques, Paris, 1870.

[Ki1] W. Killing, Die Zusammensetzung der stetigen endlichen Transformationsgruppen, erster Theil, *Math. Ann.*, 31 (1887) 252–290; zweiter Theil, 33 (1888) 1–48; dritter Theil, 34 (1888) 57–122; vierter Theil, 36 (1889) 161–189.

[Ki2] ______, Einführung in die Grundlagen der Geometrie, vol. 1, Paderborn, 1893, vol. 2, Paderborn, 1898.

[K1] F. Klein, Vergleichende Betrachtungen über neue geometrische Forschungen, Program, Erlangen, 1872 (transl. M. W. Haskell, *Bull. N.Y. Math. Soc.*, 2 (1893)).

[K2] ______, Über binäre Formen mit linearen Transformationen in sich, *Math. Ann.*, 9 (1875) 183–208.

[K3] ______, Zur nichteuklidischen Geometrie, *Math. Ann.*, 37 (1890) 544–572.

[La] E. Laguerre, Sur la théorie des foyers, *Nouv. Ann. Math.*, 12 (1853) 57–66, Oeuvres 2, 6–15.

[M] W. Magnus, Non Euclidean Tesselations and Their Groups, Academic Press, N.Y., 1974.

[Pc1] H. Poincaré, Sur les fonctions uniformes qui se reproduisent par des substitutions linéaires, *Math. Ann.* 19 (1881) 553.

[Pc2] ______, Théorie des fonctions fuchsiennes, *Acta Math.*, 1 (1882) 1–62.

[P] J. V. Poncelet, Traité des propriétés projectives des figures, Paris, 1822.

[R] B. Riemann, Über die Hypothesen, die der Geometrie zugrunde liegen, *Abh. Ges. Göttingen*, 13 (1868) 1–20; Werke, 272–287.

[S] A. Schönfliess, Krystallsysteme und Krystallstruktur, Leipzig, 1891.

[T] C. Thaer, Euklid, Die Elemente, Buch I-XIII; Ostwald's Klassiker, 235, 236, 240, 241, 243 (1933–1937), reprint Wiss. Buchgesell., Darmstadt 1962.

[vS] K. G. C. von Staudt, Geometrie der Lage, Nürnberg, 1847.

*Mathematical Cryptology for Computer Scientists and Mathematicians.* By Wayne Patterson, Rowman & Littlefield, 1987. xxii + 312 pp.

*A Course in Number Theory and Cryptography.* By Neal Koblitz, Springer-Verlag, Graduate Texts in Mathematics # 114, 1987. ii + 208 pp.

ALAN G. KONHEIM
*Department of Computer Science, University of California, Santa Barbara, CA 93106*

The books by Patterson and Koblitz are the latest entries in a growing literature on cryptography. The widening use of information processing in consumer transactions and the issue of privacy has motivated the inclusion of cryptography in undergraduate curricula. Cryptography (or cryptology) is related to various areas of computer science, including complexity theory, operating systems and data base theory and depends upon several branches of mathematics including number theory and algebra.

These two books are directed to different audiences and assume different levels of mathematical maturity.

**Mathematical Cryptology for Computer Scientists and Mathematicians.** Notwithstanding its title, this book is addressed to the computer scientist. Its forte is the first comprehensive exposition of developments within the past six years in the area of public key cryptosystems.

The first chapter of seventeen pages is devoted to *all* of cryptography before 1970! The Caesar, Vigenère, Playfair, and columnar transpositions are introduced. No discussion is given of the Hagelin machine (and its children), the Japanese 'rainbow' family, the Hebern rotor system, or additive stream cipher systems which generate a sequence $\bar{k} = (k_0, k_1, \dots)$ with $k_i \in \mathbf{Z}_2 = \{0, 1\}$ which is exclusive-ORed to plaintext $\bar{x}$ to obtain ciphertext $\bar{y} = \bar{x} \oplus \bar{k}$.

In describing the computational complexity of cryptographic systems, the author states on page 6 that "until recent years, computational infeasibility meant being too difficult to compute with pen and pencil." This is certainly an exaggeration. The cryptanalysis of the Enigma system during World War II led to the 'invention' of the digital computer and cryptography has been a constant stimulation in the development of faster and more efficient processing systems.

Chapter 2 describes the Data Encryption Standard (DES). There is no discussion of the modes of operation of DES—chaining and stream generation. A summary of published attacks on DES is given in chapter 10 in addition to the criticism of DES. As contrasted with the knapsack cryptosystem for which an attack is *known* to succeed, no cryptanalysis of DES has been published which recovers (even *part of*) the key and/or plaintext from DES-enciphered ciphertext.

Chapter 3 introduces public key cryptosystems which have altered the direction of research in cryptography. The chapter begins with a discussion of *key management*, the raison d'être of public key cryptography. Chapters 4 and 5 discuss the knapsack and RSA algorithms. The strongest feature of Patterson's book is the lovely exposition of Shamir's attack on the knapsack system given in Chapter 6 together with the excellent discussion of the $L^3$-algorithm.

Chapter 7 describes additional public key systems. Only a cursory discussion is included of the public key system proposed by El Gamal in 1985 which forms the

basis of the commercial offering by CYLINK Incorporated which uses this public key cryptosystem for key exchange.

Chapter 8, entitled 'Other Security Problems,' deals with *authentication* and *digital signatures*, in my view the most important applications of cryptography in information processing systems. While public key cryptosystems 'solve' the problem of key distribution, their use often introduces new issues. When a table of 'public key values' is stored in a multiuser system, provision must be made to protect the table against unauthorized alteration. Failure to do so opens the system to various attacks. There is no discussion of these problems. Lacking is the same formalism introduced by the author in Chapter 1 where a cryptosystem is defined as a 4-tuple $\langle K, M, C, T \rangle$. What does authentication require? What is required of a system enabling users to 'sign' messages?

A nice feature of this book is the inclusion of Pascal programs for a variety of cryptosystems.

Patterson's book is easy to read and provides the best introduction to modern public key cryptosystems. Unfortunately, due to the paucity of examples, the book is probably not suitable as a textbook.

**A Course in Number Theory and Cryptography.** Neal Koblitz's book is directed to, and will only be appreciated by, a mathematically mature audience. It presents a detailed review of factorization and techniques for the solution of the logarithm problem, two problems which arise in public key cryptography.

The first two chapters offer preliminaries on number theory and finite fields.

Chapter 3 describes some elementary cryptographic systems. The idea of a public key system is introduced in the next chapter, included are the RSA exponentiation algorithm and the knapsack cryptosystem.

The strength of cryptosystems which employ exponentiation for key exchange and encipherment depend on the supposed complexity of the logarithm problem:

*Given*: $y = q^x$ (modulo $p$) with $p$ a prime and $q$ primitive relative to $p$.

*Find*: $x$.

The current state of the logarithm problem up to, but not including, the 1984 paper by Coppersmith and Odlyzko is presented.

There is little discussion of Shamir's attack on the knapsack problem or the equally interesting work of Brickell.

Chapter 5 is devoted to primality and factorization. Several public key cryptosystems require the 'selection of large' primes. A 'large' integer $n$ is to be tested for primality. The chapter describes the Solovay-Strassen and Rabin-Miller tests.

Although it has *not* been proved that the strength of the RSA encipherment algorithm is equivalent to the complexity of factorization, the evidence to date suggests this tentative conclusion is not unreasonable. The remainder of the chapter is devoted to an exposition of factorization methods.

Modern cryptography has provided the motivation for research into factorization. Some of the most exciting new work lies within algebraic geometry. The final chapter describes the recent work on elliptic cryptosystems.

Koblitz's book is well written and a welcome addition to the literature on cryptography.

*Invitation to Complex Analysis.* By Ralph Philip Boas. The Random House/Birkhäuser Mathematics Series. Random House, New York, 1987. xii + 347 pp.

GEORGE PIRANIAN
*Department of Mathematics, University of Michigan, Ann Arbor, MI 48109*

> Then, said Stephen, you pass from point to point, led by its formal lines; you apprehend it as balanced part against part within its limits; you feel the rhythm of its structure. In other words the synthesis of immediate perception is followed by the analysis of apprehension.
>
> James Joyce, *A Portrait of the Artist as a Young Man*

What is it that we call complex analysis? According to *The American Heritage Dictionary*, analysis is the separation of an intellectual or substantial whole into constituents for individual study. This definition pleases the mind, for the Greek word *αναλυειν* means to unloose, release, or dissolve.

For mathematical contexts, the same dictionary gives two special meanings of analysis: "Methodology principally involving algebra and calculus as opposed to synthetic geometry, group theory, and number theory," and "the method of proof in which a known truth is sought as a consequence of reasoning from the thing to be proved." Neither of the two descriptions will ever attain canonical status.

Boas avoids the folly of an impossible definition by making a modest declaration. In his preliminary statement to students he writes: "Complex analysis was originally developed for its applications; however, the subject now has an independent and active life of its own, with many elegant and even surprising results." The declaration does not characterize complex analysis; but complex analysts know that no reasonable description of their territory could ever have remained satisfactory for more than a quarter century.

An understanding of the term *complex analysis* requires at least peripheral participation. To those who by virtue of necessity or volition refrain from participation we might simply say that complex analysis began as the art of using complex-valued functions in the analysis of various physical problems and that today it is primarily the study, by analysis and synthesis, and with geometric, topological, algebraic, number-theoretic, or other cultural orientations, of complex-valued functions in spaces of one or more complex variables.

Even the consensus that a course in complex analysis should begin with the introduction of complex numbers allows serious differences of opinion about the first lesson. Whether teachers take a historical, heuristic, axiomatic, or constructive approach and whether they think primarily in terms of algebraic or geometric models depends on their taste. Taste in turn depends on personal predilections and experience, on passion for elegance, generality, or detail, and on degrees of resonance with the educational system's temporary philosophies. If students are fortunate, their teacher will also adjust the message to their background and maturity.

Once we know what we mean by complex numbers and by functions, neighborhoods, limits, and derivatives, we can come to grips with the concept of analytic (holomorphic) functions. At about this stage, we meet the Cauchy-Riemann equations, an illuminating example of analysis in the sense of the dictionary's primary definition: separation of significant aspects for the sake of detailed examination.

The ability to recognize instinctively what is significant sets apart the mathematicians from shufflers of formulas, just as it distinguishes creative musicians, sculptors, and writers from hacks.

Giants among our predecessors constructed a sturdy foundation for the complex function theory developed under the stimulus of ingenious maneuvers associated with practical applications. In other words, they provided a platform of concepts, definitions, and theorems that stands without swaying and twisting in the winds of shifting contexts. Naturally, there followed a period of primarily analytical investigations, and these investigations produced a large volume of elegant and astonishing information. Examples: The calculus of residues, the theorem on conformal mappings of simply connected domains, and Carathéodory's theory of prime ends.

The rise of topology influenced us to study not only the elements of various classes of functions, but also local and global structural properties of function spaces. At the same time, it became clear that the search for theorems of the form "A implies B" could not forever satisfy all our aesthetic needs. The Polish school of mathematics led the search for theorems of the form "A does not imply B." Surely, the construction of counterexamples has often depended on synthesis (putting together) rather than analysis. Despite its flaws, we can bless even the first half of the twentieth century.

To illustrate the distinction between the methods of analysis and synthesis, let us first sketch the way in which we encourage students to organize their thoughts about the global behavior of the function that carries each point $x + iy$ in the complex plane to the corresponding point $e^x(\cos y + i \sin y)$. We focus the students' attention on the horizontal strips of width $2\pi$ in the domain, study the map of each strip, fasten the maps of adjoining strips to each other along slits from 0 to the point at infinity, and offer the Riemann surface thus obtained both as a geometric toy and as a mnemonic device. By way of contrast to this basically analytical process, we recall the late Gerald MacLane's and his friends' fabrications of outrageous analytic functions in terms of Riemann surfaces. The goal was not the analysis of an object that had long been dear to the mathematical culture. Rather, working with a stack of copies of the unit disk or the complex plane, a pair of scissors, and a pot of glue, the artist created an ingenious monstrosity. Because his beast was simply connected and had at least two boundary points, it was the Riemann surface of some analytic function in the unit disk, and, *voilà*, everybody could see that analytic functions with certain unexpected properties exist.

Let us now turn to the book. Boas obviously believes in exercises, and he has confidence in the soundness of his pedagogical opinions. For example, he objects to the use of mathematical symbols where words can deliver a clearer message: on page 3, he introduces the stereographic projection in a simple paragraph supported by a drawing but devoid of formulas, and already on page 14 we meet Landau's $o$- and $O$-notation, one of the most effective devices for the elimination of computational pedantry. Section 7H (pages 64–68) involves substantial manipulations; but in return the student becomes acquainted with the Gamma and Beta functions. Boas is a master at conveying central truth without telling everything he knows.

Although the book's title does not promise a sit-down dinner, that's what we get. An important aspect of the occasion is that no curriculum committee has caterer's rights. The MONTHLY's editorial rules forbid recitations from the menu; but I can report that the host has chosen and prepared the food in his own inimitable way.

The meal's nutritional balance is above reproach, and each dish presents itself attractively. The herbs come from Boas's private garden, and the beverages have aged in his cellar.

*Principles of Computer Science.* By M. Sandra Carberry, A. Toni Cohen, and Hatem M. Khalil. Computer Science Press, 1986, xvi + 636 pp.

STANLEY E. SELTZER
*Department of Mathematics and Computer Science, Ithaca College, Ithaca, New York 14850*

"What is computer science?" While this question can prompt sharp debate among those who are generally considered to be computer scientists, it is at the heart of any discussion of a computer science curriculum. Thus, any book with a title such as "Principles of Computer Science" has the potential to provoke interest. The authors of this text seem to have opinions as evidenced by their subtitle, "Concepts, Algorithms, Data Structures, and Applications" although, after raising the question on the first page of the text, they delicately evade this issue with the statement, "Fortunately, the fact that we cannot define computer science doesn't keep us from learning about or studying all facets of the discipline."

There is also the related question, "What is appropriate content for an introductory computer science course?" Here, too, there are a variety of responses. Such a course, dubbed "CS1," has been included in every computer science curriculum since Curriculum '68 [**2**]. It is probably fair to say that the trend is from a "computer programming" course to a "computer science" course, the former concentrating most heavily on writing computer programs and the latter offering what is often called a "disciplined approach to program development," that is, one that explicitly emphasizes problem solving and/or programming methodology. It should also be noted that there are individuals who advocate a nonprogramming introductory course.

The evolution from a course oriented heavily toward programming to a broader approach is quite evident in textbooks. Many recent texts begin with several chapters dealing with such topics as perspective on computers, the program development process, and problems and problem solutions. Titles, too, seem to have changed: a programming language (most commonly Pascal) was frequently the focus of the title (for example, *Oh! Pascal!* [**1**]) while newer titles, even those which stress programming, such as *Introduction to Computers and Programming: Pascal* [**5**] and *Programming by Design: A First Course in Structured Programming* [**4**], deemphasize or do not mention a specific language. The book under review may be a first, its title referring explicitly to neither programming nor to a programming language. However, before returning to these issues, it is useful to examine the book.

Part I, "Foundations," begins with a short "what is computer science" chapter but is actually a relatively detailed introduction to algorithms. There is a chapter devoted to complexity and another on verification. Certainly the inclusion of these topics is not revolutionary, nor is this particular treatment, but placement of this material so early in a text is unusual.

"Systems and Languages" is the title of Part II. Some of the material here is fairly standard: computer representation of data and system software (assemblers,

compilers, and operating systems), for example. In addition, Backus-Naur Form and parsing are introduced, and there is a chapter on computer architecture, featuring a hypothetical machine, discussion of numeric and symbolic coding, and several pages (labelled "optional") on addressing modes.

The "traditional" programming material is in Part III entitled "Programming with Pascal." These six chapters comprise a bit over one third of the text. Here the presentation is quite conventional. Although the book begins with a strong emphasis on algorithms, the presentation is rather closely tied to Pascal, and much of the exposition is of the form Problem-Solution, where Solution consists solely of a Pascal program.

The title of Part IV is "Applications." Its first chapter is a standard CS1 treatment of sorting and searching. Files are introduced in the "Information Systems" chapter, but discussion here goes beyond Pascal to mention direct access and indexed sequential files and database management systems. Further applications are scientific (including Gaussian elimination, Monte Carlo methods, and some discussion of errors of computation), artificial intelligence (with sections on expert systems and natural language processing), and software engineering. The "Perspectives" section addresses two topics, impact of computers on society and future directions, which are more commonly associated with a computer literacy course than with CS1.

What kind of CS1 course does the present text suggest? In terms of content, the book is quite ambitious. Clearly intended for a single-semester course, it contains more than enough material. This raises the usual dilemma for the instructor—what to include and what to omit. Since the distinctive aspect of this text is its approach, presumably one would not give short shrift to the foundations. This choice would seem to impart a somewhat theoretical flavor, with complexity and verification of algorithms preceding actual programming.

The mathematical level of the presentation is uneven. In the treatment of analysis of algorithms, "big-oh" notation is introduced; several examples are worked out; and an ordered list of standard functions is given. Unfortunately, no general results are stated (like the order of a polynomial is the order of the leading term) and a few pages later there is a table (not a graph) illustrating constant, log $N$, $N$, $N^2$ and $2^N$ functions. The treatment of converting from one base to another does not mention modular arithmetic, referring simply to the remainder; but the section on linear congruential random number generators does.

In short, the approach to CS1 in *Principles of Computer Science* is different from many current texts. It does attempt to address diverse aspects of computer science, rather than merely to present a "disciplined" approach to programming.

Returning to the question posed at the beginning of this essay, we note Peter Denning's article [3] entitled "What is Computer Science?" Addressing himself to a broad range of scientists, Denning, editor in chief of the Association for Computing Machinery's flagship publication, *Communications of the ACM*, lists eleven areas which have "made the transition from poorly understood sets of techniques to well-understood sets of core principles" of computer science. These are, in chronological order: theory, numerical computation, architecture, programming languages and methodology, algorithms and data structures, operating systems, networks, human interface, database systems, concurrent computation, and artificial intelligence. Except for human interface, all are dealt with to some extent by Carberry,

Cohen, and Khalil. Thus, their text attempts to take CS1 a step beyond many of its predecessors. If one may characterize the previous approach as asserting that programming is more than coding, *Principles of Computer Science* insists that computer science is more than programming.

REFERENCES

1. Doug Cooper and Michael Clancy, Oh! Pascal!, second edition, W. W. Norton & Co., 1985.
2. Curriculum Committee on Computer Science, Curriculum '68—Recommendations for the undergraduate program in computer science, *Communications of the ACM* 11 (1968) 151–197.
3. Peter Denning, What is computer science?, *American Scientist*, 73 (1985) 16–19.
4. Phillip L. Miller and Lee W. Miller with Purvis M. Jackson, Programming by Design: A First Course in Structured Programming, Wadsworth Publishing Company, 1987.
5. Peter P. Smith, Introduction to Computers and Programming: Pascal, Wadsworth Publishing Company, 1987.

*Writing Mathematics Well.* By Leonard Gillman. The Mathematical Association of America, Washington, D.C., 1988. ix + 49 pp.

PETER D. LAX
*NYU-Courant Institute of Mathematical Sciences*

Gillman's little pamphlet should be required reading for *every candidate for a Ph.D. in mathematics.* It is full of practical advice, easy to follow, ignored at one's peril. Although some of the admonitions are matters of common sense, Gillman escapes sounding like Polonius by virtue of a breezy style and a light touch. Many passages bring to mind *The Elements of Style* by William Strunk, Jr., and E. B. White.

All aspects of writing a mathematical paper are commented on. Section 2 gives a sound description of the principles and practices of organizing a paper. A brief paragraph on paragraphing would be a welcome addition here. Sections 4 and 5 on Mathematical English and symbols are excellent, with a nice example of what to avoid and how. Section 6 is a crisp dissertation on English usage, with a little Latin and less Greek thrown in. The author shows us how to form the plural of words like "criterion," but not of "lemma" (lemmata).

I would have welcomed a stirring call to restore *Q.E.D.* to its rightful place at the end of proofs; it is crisp, sanctified by long tradition, and has a nice doomsday cadence.

Section 8 on the mechanics of preparing a manuscript and seeing it through to print, excellent for now, is likely to become obsolete because of changes in technology. For instance, faxing may be an efficient way of returning corrected galleys.

Section 3, on presenting results, has the most mathematical substance, and here I have some substantial comments:

In subsection 3.8 Gillman observes that proofs by contradiction can be classified as *essential*, *questionable*, and *spurious*; *questionable* means that the indirection merely amounts to replacing a proposition by its contrapositive. He cites the classical proof of the irrationality of $\sqrt{2}$ as a true-blue, essentially indirect proof. I

disagree, for the following reason:

The nub of the statement is that *there is no rational number whose square is* 2. This is the same as saying that the square of any rational number is something other than 2. To see this write $r$ in reduced form as $p/q$. Then

$$r^2 = \frac{p^2}{q^2}$$

is also in reduced form. The reduced form of 2 is $2/1$; since $p$ is never 2, neither is $r^2$. *Q.E.D.*

For me, a prototype indirect proof is the one used by Friedrichs [**2**] to show that the range of a positive differential operator $P$ includes all of $L^2$. As a first step he shows that the range of $P$ is dense in $L^2$; for, if not, there would exist a nonzero $f$ in $L^2$ that is orthogonal to the range of $P$. Such an $f$ would belong to the nullspace of $P^*$, the adjoint of $P$. By a separate argument he shows that the nullspace of $P^*$ contains no nonzero functions.

Pólya has cautioned his students about indirect proofs, not because he was squeamish about invoking "tertium non datur," but because he found it distasteful to dwell in a world containing a contradiction.

All the examples cited by Gillman come from pure, abstract mathematics. His implied paradigm for a mathematical paper is basically Landau's *Definition*, *Satz*, *Beweis*, although fleshed out and provided with a human face. The presentation of results in applied mathematics calls for different strategies, not addressed in this pamphlet. These are illustrated in such outstanding expositions of applied mathematics as Mark Kac's masterly Carus Monograph [**3**], recently reprinted, or Barry Cipra's excellent introduction to the Ising Model [**1**].

The pamphlet under review is twice blessed: the reader learns from what the author says, and how he says it.

REFERENCES

1. Barry A. Cipra, An introduction to the Ising model, *American Mathematical Monthly*, 94 (1987) 937–959.
2. K. O. Friedrichs, The identity of weak and strong extensions of differential operators. *Trans. AMS*, 55 (1944) 132–151.
3. M. Kac, Statistical Independence in Probability Analysis and Number Theory, Carus Monograph, MAA, 1959.

## P. G. O. M.*

Who can take prime numbers and find their density
Using only techniques he calls elementary?
P.G.O.M. can show it's one over log $n$.
P.G.O.M. can, he just mumbles log log $n$
To make it all come true!

Who can take each edge with fixed probability
Put them all together and prove connectivity?
P.G.O.M. can do it asymptotically,
P.G.O.M. can, he just mumbles log log $n$
To make it all come true!

Who'll tell you that his age
Is more than you would gauge?
By billions you would slight the figure.
Although it can't be found with rigor,
We're glad to say its growing bigger.

Couples in his lingo are known as boss and slave.
Put the two together, that's how epsilons are made,
In Uncle Paul's slang, oh in Uncle Paul's slang.
Raise your glasses in a toast, drink the poison you like most
And he'll respond with O.J.

Who often gets to look
Into God's perfect book?
Proofs chosen by The Great Librarian
Are elegant and slick, not hairy, an'
They often sound a bit Hungarian!

Who can take a sequence, partition it at will,
Find arithmetic progressions remaining in it still?
P.G.O.M. can, by citing Van der Waerden,
P.G.O.M. can, he just mumbles log log $n$
To make it all come true!

Who else can help you take
An Erdös number great,
Collaborate with skill and power,
Reduce it to one in an hour,
And have time left to take a shower?

Whose mind is always open to problems old and new,
And who makes his conjectures worth a thousand bucks or two?
Uncle Paul can, and he seldom has to pay.
Just try to solve a few, you'll find out they're hard to do,
It always happens that way.

And yet, P.G.O.M. can, he just mumbles log log $n$
To make it all come true.

---

*Poor Grand Old Man, to the tune of "The Candy Man."
Lyrics by Allen J. Schwenk, Western Michigan University, Kalamazoo.
Presented to Paul Erdös in celebration of his 75th birthday at the Sixth International Conference on the Theory and Applications of Graphs, in Kalamazoo, 2 June 1988.

# TELEGRAPHIC REVIEWS

*Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.*

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

| | | | | | |
|---|---|---|---|---|---|
| T: | Textbook | P: | Professional Reading | 1–4: | Semesters |
| C: | Computer Software | L: | Undergraduate Library | **: | Special Emphasis |
| S: | Supplementary Reading | 13: | Grade Level | ??: | Questionable |

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

**General, S*, L**.** *Innumeracy: Mathematical Illiteracy and Its Consequences.* John Allen Paulos. Hill & Wang, 1988, 135 pp, $16.95. [ISBN: 0-8090-7447-8] An outspoken, often angry exposé of public obliviousness to quantitative meanings, from national debt to risks of terrorism, from AIDS testing to the ubiquity of coincidences. Paulos links innumeracy with the alarming spread of pseudoscience, which he claims is fueled by attributions of cause and effect to incidents that are purely coincidental. LAS

**Foundations, P, L.** *Surfaces.* Avrum Stroll. U of Minnesota Pr, 1988, xi + 227 pp, $39.50; $15.95 (P). [ISBN: 0-8166-1693-0; 0-8166-1694-9] Not about the geometry of surfaces but about their *philosophy:* what they are (abstractions or reality), how they relate (to perception, to faces, to boundaries), and what they mean. A work in the foundations of empirical (not axiomatic) geometry with possible implications for perception and visualization studies in artificial intelligence. LAS

**Foundations, P*, L*.** *The Universal Turing Machine: A Half-Century Survey.* Ed: Rolf Herken. Oxford U Pr, 1988, xiv + 661 pp, $115. [ISBN: 0-19-853741-7] A commemorative volume written a half-century after Turing's 1937 "Entscheidungsproblem" paper. Five expositions of the context of Turing's work are followed by 23 diverse examples of current research—from biology to physics, from logic to number theory—that follows directly from Turing's ideas. LAS

**Combinatorics, P.** *Algebraic, Extremal and Metric Combinatorics, 1986.* Ed: M-M. Deza, P. Frankl, I.G. Rosenberg. London Math. Soc. Lect. Note Ser., V. 131. Cambridge U Pr, 1988, ix + 245 pp, $34.50 (P). [ISBN: 0-521-35923-6] An apparently—but not actually—disparate collection of papers presented at the 1986 Montreal Conference on Algebraic Combinatorics and Extremal Problems. Most interesting is the intertwining of apparently distant parts of combinatorial theory. SS

**Number Theory, P, L.** *Mathematische Werke, Band I and II, Second Edition.* Gotthold Eisenstein. Chelsea, 1989, $120 set [0-8284-1280-X]. *Band I,* xxiii + 502 pp; *Band II,* xiii + 437 pp. Contains, in addition to what is in the *First Edition* (TR, March 1976), an essay by André Weil on Eisenstein's life and work and another short autobiography. JD-B

**Number Theory, P.** *New Advances in Transcendence Theory.* Ed: Alan Baker. Cambridge U Pr, 1988, xii + 434 pp, $79.50. [ISBN: 0-521-33545-0] Proceedings of the July 1986 symposium held at the University of Durham. Topics range from linear forms in logarithms to Diophantine equations to transcendence theory of classical functions, written by experts in the field. A sequel to the ten-year-old *Transcendence Theory: Advances and Applications.* GG

**Number Theory, S(15-16).** *Exploring Number Theory with Microcomputers.* Donald D. Spencer. Camelot, 1989, 288 pp, $19.95 (P). [ISBN: 0-89218-113-3] The author attempts to "introduce the reader to computer programming using number theory examples." Included are 85 BASIC programs on various topics in elementary number theory. SG

**Linear Algebra, T(15-16), S, L.** *Exercices d'Algèbre, Linéaire et Bilinéaire.* J.-B. Hiriart-Urruty, Y. Plusquellec. CEPADUES-Editions, 1988, 307 pp, (P). [ISBN: 2-85428-187-X] An interesting collection of problems, many with hints and all with at least sketches of solutions. Assumes knowledge of elementary linear algebra and considerable mathematical maturity. JD-B

**Group Theory, T(15-16: 1), S, L.** *Groups and Symmetry.* M.A. Armstrong. Undergrad. Texts in Math. Springer-Verlag, 1988, xi + 186 pp, $34. [ISBN: 0-387-96675-7] Introduction to group theory, with emphasis on examples, especially of symmetry groups. Covers standard topics such as Langranges' theorem, the Sylow theorems, and the classification of finitely generated Abelian groups as well

as matrix groups, lattice groups, wallpaper patterns, and a proof of the Nelson-Schreier theorem using trees. A fun approach to groups. LC

**Algebra, T(18), S, P.** *Buildings.* Kenneth S. Brown. Springer-Verlag, 1989, xiii + 215 pp, $39. [ISBN: 0-387-96876-8] An interesting and engaging introduction to Tits's theory of buildings and its application to group theory. Loaded with insightful motivating remarks, historical comments, and references to the literature. SG

**Partial Differential Equations, S(18), P.** *Lecture Notes in Mathematics-1341: Elliptic Boundary Value Problems on Corner Domains.* Monique Dauge. Springer-Verlag, 1988, viii + 259 pp, $24.30 (P). [ISBN: 0-387-50169-X] Develops and applies theory of boundary value problems—especially the Dirichlet problem—on domains with various kinds of singularities: corners, edges, cracks, holes, etc. The general setting is Sobolev spaces; one of four appendices collects pertinent Sobolev theory. PZ

**Partial Differential Equations, T(18: 1, 2), S, P.** *Methods Based on the Wiener-Hopf Technique for the Solution of Partial Differential Equations.* B. Noble. Chelsea, 1988, x + 246 pp, $19.95. [ISBN: 0-8284-0332-5] A textually unaltered reprint of the 1958 original version. In the Wiener-Hopf technique, as modified by D.S. Jones, the complex Fourier transform is applied to a given partial differential equation to yield a complex variable equation. The latter is solved by analytic continuation; complex Fourier inversion is applied to the result. Expository level appropriate for graduate students; with many worked examples. PZ

**Numerical Analysis, P.** *Finite Elements: Theory and Application.* Ed: D.L. Dwoyer, M.Y. Hussaini, R.G. Voigt. ICASE/NASA LaRC Series. Springer-Verlag, 1988, x + 302 pp, $48.50. [ISBN: 0-387-96610-2] Proceedings of a workshop held July 28-30, 1986 in Hampton, Virginia. LC

**Numerical Analysis, S(16-17), P, L**.** *Handbook for Matrix Computations.* Thomas F. Coleman, Charles Van Loan. Frontiers in Appl. Math. SIAM, 1988, vii + 264 pp, $24 (P). [ISBN: 0-89871-227-0] Designed to be used as a reference by those needing to do scientific computations using matrices. After a brief introduction to Fortran 77, it covers BLAS (a standard collection of Basic Linear Algebra Subprograms), LINPACK (a widely-used collection of Fortran codes for many linear equations and least squares calculations), and MATLAB (an interactive system for matrix computation). AO

**Numerical Analysis, T(16-18: 1, 2), S, P, L.** *Numerical Methods for Initial Value Problems in Ordinary Differential Equations.* Simeon Ola Fatunla. Comput. Sci. & Sci. Comput. Academic Pr, 1987, xi + 295 pp, $44.50. [ISBN: 0-12-249930-1] "This text will be essentially dedicated to the theory, development, and implementation of both currently existing and new algorithms based on the discrete variable approach." The first half is suitable for advanced undergraduates, while the latter half is intended for graduates and researchers. Includes recent results in the numerical treatment of singular and discontinuous initial value problems and ordinary differential equation solvers, and suggests possible research areas. LC

**Numerical Analysis, P.** *Difference Methods for Initial-Boundary-Value Problems and Flow Around Bodies.* Zhu You-lan, *et al.* Springer-Verlag, 1988, viii + 600 pp, $120. [ISBN: 0-387-10887-4] The 1980 Science Press Beijing volume. Concerns itself with revision of numerical methods for discontinuous solutions of quasi-linear hyperbolic systems of partial differential equations. JAS

**Functional Analysis, P.** *Lecture Notes in Mathematics-1332: Functional Analysis.* Ed: E. Odell, H. Rosenthal. Springer-Verlag, 1988, 202 pp, $20 (P). [ISBN: 0-387-50018-9] Proceedings of a seminar in functional analysis held at the University of Texas at Austin, 1986-87. LC

**Functional Analysis, T(16-17: 1, 2), L.** *An Introduction to Hilbert Space.* Nicholas Young. Cambridge U Pr, 1988, x + 239 pp, $59.50; $21.95 (P). [ISBN: 0-521-33071-8; 0-521-33717-8] An introductory textbook in Hilbert space theory for honors undergraduates and graduate students, with an eye towards applications, particularly to solution of partial differential equations in mathematical physics, and to approximation of functions in complex analysis. Presumes introductory real analysis, linear algebra, metric spaces. Over two-hundred problems, about fifty with answers. RB

**Functional Analysis, T(18: 1, 2), S, P.** *Amenability.* Alan L.T. Paterson. Math. Surveys & Mono., No. 29. AMS, 1988, xvii + 452 pp, $90. [ISBN: 0-8218-1529-6] A group or a semi-group $G$ is called amenable if there exists a left-invariant mean $m$ on $G$. While the notion of amenability is used mainly to analyze locally-compact groups, it also has applications in statistics, differential geometry, and operator algebra. The author writes: "The ubiquity of amenability ideas and the depth of the mathematics with which the subject is involved seems evidence to the author that here we have a topic of fundamental importance in modern mathematics, one that deserves to be more widely known than it is at present." A knowledge of abstract harmonic analysis and of functional analysis are necessary to read this book. Problems follow each chapter. LW

**Analysis, P.** *The Finite Calculus Associated with Bessel Functions.* Frank M. Cholewinski. Contemp. Math., V. 75. AMS, 1988, xi + 122 pp, $19 (P). [ISBN: 0-8218-5083-0] Develops generalized umbral calculus associated with the Euler operator. LC

**Analysis, P, L.** *Multivariate Splines.* Charles K. Chui. CBMS-NSF Reg. Conf. Ser. in Appl. Math., V. 54. SIAM, 1988, vi + 189 pp, $19 (P). [ISBN: 0-89871-226-2] Based on lectures presented at the Regional Conference on "Theory and Applications of Multivariate Splines" held at Howard University, August 10-14, 1987. An elementary introduction that parallels the more familiar theory of univariate

splines. Both theoretical results and computational methods are presented. AO

**Analysis, S(18), P.** *Linear Spaces and Differentation Theory.* Alfred Frölicher, Andreas Kriegl. Wiley, 1988, xvi + 246 pp, $67.95. [ISBN: 0-471-91786-9] Desiring a differentiation theory in infinite-dimensional and non-normable linear spaces, the authors study a class of linear spaces which includes Frechet spaces. This class of "convenient" vector spaces allows the standard categorical constructions, yet it also has the internal properties needed for analysis. Beginning with an introduction to some category theory, the book proceeds to develop calculus in convenient vector spaces, then to study closedness properties and various topologies on these spaces. LW

**Analysis, P.** *Educational Computing in Mathematics: ECM/87.* Ed: T.F. Banchoff, *et al.* North-Holland (US Distr: Elsevier Science), 1988, ix + 284 pp, $79. [ISBN: 0-444-70456-6] Twenty-four papers from the June 1987 International Congress on Educational Computing in Mathematics held at Rome. Although all refer to education, half the papers primarily concern computing as an experimental tool in mathematics research. The other half stress pedagogical, theoretical, and practical issues. Brief software reports are collected in a final section. A useful reference and source of examples, especially for computing at the advanced undergraduate level. PZ

**Analysis, P.** *Nonlinear Problems in Abstract Cones.* Dajun Guo, V. Lakshmikantham. Notes and Reports in Math. in Sci. & Eng., V. 5. Academic Pr, 1988, viii + 275 pp, $34.95. [ISBN: 0-12-293475-X] Discusses solutions to nonlinear integral and differential equations in abstract cones (closed convex subset of a real Banach space). Begins with basic properties of cones, then addresses positive fixed points of some nonlinear operators. LC

**Algebraic Geometry, P.** *Introduction to Arakelov Theory.* Serge Lang. Springer-Verlag, 1988, x + 187 pp, $49.95. [ISBN: 0-387-96793-1] Transposes results of arithmetical algebraic geometry to number fields by including Archimidean places as points. Proves adjunction formula, Hodge Index Theorm, Faltings Riemann-Roch Theorem for curves over rings of integers of number fields. Appendix includes conjectures and recent work. SB

**Algebraic Geometry, P.** *The Curves Seminar at Queen's, Volume V.* Ed: Anthony V. Geramita. Papers in Pure & Appl. Math., No. 80. Queen's U, 1988, 102 pp, (P). A collection of five expository papers by Ramanan, Robbiano, Yuzvinsky, Geramita and Migliore, and Geramita and Valabrega. SG

**Differential Geometry, P.** *Lecture Notes in Mathematics-1314: Kähler-Einstein Metrics and Integral Invariants.* Akito Futaki. Springer-Verlag, 1988, iv + 140 pp, $13.90 (P). [ISBN: 0-387-19250-6] A study of the existence of Kähler-Einstein metrics on compact complex manifolds of positive first Chern class. JAS

**Geometry, P.** *Lecture Notes in Mathematics-1346: Topology and Geometry—Rohlin Seminar.* Ed: O. Ya. Viro. Springer-Verlag, 1988, xi + 581 pp, $52.60 (P). [ISBN: 0-387-50237-8] A range of papers, most presented in the V.A. Rohlin Seminar from 1984-1986, collected in honor of V.A. Rohlin. JAS

**Geometry, T(16-18: 1), S, P.** *Automorphisms of Surfaces after Nielsen and Thurston.* Andrew J. Casson, Steven A. Bleiler. London Math. Soc. Student Texts, V. 9. Cambridge U Pr, 1988, 105 pp, $34.50; $12.95 (P). [ISBN: 0-521-34203-1; 0-521-34985-0] Lecture notes on a course given by the first author. An elementary, but very (too?) concise presentation of some lovely geometry. Assumes introductory topology and linear algebra and a significant amount of mathematical maturity and sophistication. Especially good for concrete geometric ideas. JAS

**Algebraic Topology, P.** *Lecture Notes in Mathematics-1323: Boundedly Controlled Topology.* Douglas R. Anderson, Hans J. Munkholm. Springer-Verlag, 1988, xii + 309 pp, $28.60 (P). [ISBN: 0-387-19397-9] An exposition of the algebraic topology of categories of boundedly controlled topological spaces with special emphasis on cobordism. JAS

**Operations Research, T(16-18: 1), L.** *The Theory of Games.* Wang Jianhua. Oxford Math. Mono. Clarendon Pr, 1988, viii + 162 pp, $53. [ISBN: 0-19-853560-0] A concise treatment of game theory, ranging from fundamental concepts and properties to some recent results. Topics include: matrix games, continuous games, $n$-person non-cooperative games, and $n$-person cooperative games. RH

**Operations Research, P.** *Search Theory: Some Recent Developments.* Ed: David V. Chudnovsky, Gregory V. Chudnovsky. Lect. Notes in Pure & Appl. Math., V. 112. Marcel Dekker, 1989, xi + 155 pp, $89.75 (P). [ISBN: 0-8247-8000-0] Search theory is an important part of operations research which focuses on problems of optimization, game theory, differential games, and statistics. This text summarizes the latest developments in search theory. Includes classical, differential equations, optimal control, game-theoretic, and statistical and ergodic theory approaches. RH

**Optimization, P.** *Lecture Notes in Control and Information Sciences-113: System Modelling and Optimization.* Ed: M. Iri, K. Yajima. Springer-Verlag, 1988, x + 787 pp, $120 (P). [ISBN: 0-387-19238-7] Seventy-three contributed papers plus four papers from the plenary sessions from the Thirteenth IFIP Conference on System Modelling and Optimization, Tokyo, Japan, August 31-September 4, 1987. LC

**Optimization, T(17-18), P, L.** *Fractional Programming.* B.D. Craven. Sigma Ser. in Appl. Math., V. 4. Heldermann Verlag, 1988, iii + 145 pp, $29 (P). [ISBN: 3-88538-404-3] Covers the theory and applications of fractional programming—optimization problems where the objective function is a quotient. Discusses linear and nonlinear cases, duality and sensitivity, and presents some algorithms. Text type is dot matrix. LC

**Optimization, P.** *New Computer Methods for Global Optimization.* H. Ratschek, J. Rokne. Math. & Its Applic. Halsted Pr, 1988, iii + 229 pp, $49.95. [ISBN: 0-470-21208-X] Presents methods based on interval arithmetic for computing *global* solutions to optimization problems. Particular attention is given to combining standard algorithms for nonlinear optimization with the techniques introduced in this book. Both constrained and unconstrained problems are considered. AO

**Dynamical Systems, S**(17).** *Dynamical Systems I: Ordinary Differential Equations and Smooth Dynamical Systems.* Ed: D.V. Anosov, V.I. Arnold. Ency. of Math. Sci., V. 1. Springer-Verlag, 1988, ix + 233 pp, $59. [ISBN: 0-387-17000-6] Two books in one. The first, by Arnold and Il'yashenko, is on the local theory of ordinary differential equations with much attention focused on the analysis of singularities. Does not include bifurcation theory. The second, by Anosov, *et al*, is on smooth (as opposed to topological or measurable) dynamical systems. Topics covered include topological dynamics and flows on surfaces. Together, these provide an accessible introduction to dynamical systems. The style is simple, the authors use many examples motivating definitions and theorems. No exercises. MR

**Dynamical Systems, T*(15-16: 1), S, P, L*.** *Fractals Everywhere.* Michael Barnsley. Academic Pr, 1988, xii + 394 pp, $39.95. [ISBN: 0-12-079062-9] The first textbook of fractal geometry, based on a course at Georgia Tech. Introduces lots of good mathematics (metric spaces, affine transformations, contraction mappings, measures) in the context of fractal models of natural images. Central focus is on compression and reconstruction of images that are too complex to be encoded efficiently point-by-point, but that can be approximated by algorithmic transformations that describe iterative relations among self-similar parts. Prerequisite: calculus and linear algebra. LAS

**Dynamical Systems, P.** *Lecture Notes in Mathematics-1342: Dynamical Systems.* Ed: J.C. Alexander. Springer-Verlag, 1988, viii + 725 pp, $61.70 (P). [ISBN: 0-387-50174-6] Proceedings of the special year in dynamical systems held at the University of Maryland during 1986-87, which consisted in part of three separate conferences: Ergodic Theory and Topological Dynamics; Symbolic Dynamics and Coding Theory; and Smooth Dynamics, Dynamics, and Applied Dynamics. Among the 38 research papers presented here is an earlier but notable joint paper by J. Milnor and W. Thurston on iterated maps of the interval, which circulated widely in preprint form. CE

**Control Theory, P, L*.** *Report of the Panel on Future Directions in Control Theory: A Mathematical Perspective.* Wendell H. Fleming. SIAM, 1988, 98 pp, (P). [ISBN: 0-89871-234-3] Survey of modern control theory—modeling, history, applications—with analysis of research opportunities, educational needs, and professional issues. Eleven vignettes provide specific illustrations: stochastic, adaptive, optimal, and robust control; nonlinear and linear systems; robotics; variational problems; and more. Report of a special panel formed to assess the status and needs of this particular field. Provides a good overview, especially for non-experts. LAS

**Stochastic Processes, P.** *Multidimensional Brownian Excursions and Potential Theory.* K. Burdzy. Pitman Res. Notes in Math. Ser., V. 164. Longman Scientific & Technical (US Distr: Wiley), 1987, 172 pp, $44.95 (P). [ISBN: 0-470-20892-9] A Brownian excursion is that part of a trajectory of Brownian motion which lies in a given set. The book presumes substantial background and attempts to pull together state-of-the-art results. TAV

**Stochastic Processes, P.** *Stochastic Processes in Physics and Engineering.* Ed: Sergio Albeverio, *et al.* Math. & Its Applic. D Reidel (US Distr: Kluwer Academic), 1988, xi + 415 pp, $89. [ISBN: 90-277-2659-0] Proceedings of a 1986 meeting at Bielefeld. Twenty papers photocopied and collected with a useful index. TAV

**Computational Statistics, P, L.** *Principles of Random Variate Generation.* John Dagpunar. Clarendon Pr, 1988, xvii + 228 pp, $69. [ISBN: 0-19-852202-9] Methods and algorithms for generating random variates. Begins with the uniform distribution and general methods for generating non-uniform from uniform variates. Discusses general and distribution-specific methods for common distributions (continuous, discrete, and multivariate), with algorithms and recommendations based on speed and programming effort. Good resource for simulation writers. TH

**Statistics, P.** *Properties of Estimators for the Gamma Distribution.* K.O. Bowman, L.R. Shenton. Stat: Textbooks & Mono., V. 89. Marcel Dekker, 1988, xvi + 268 pp, $65. [ISBN: 0-8247-7556-2]

**Statistics, T(18: 2).** *Principles of Multivariate Analysis: A User's Perspective.* W.J. Krzanowski. Oxford Stat. Sci. Ser., V. 3. Clarendon Pr, 1988, xxi + 563 pp, $135. [ISBN: 0-19-852211-8] Comprehensive introduction to the principles of practical multivariate analysis. Designed for the user of multivariate methods who has a strong background in matrix algebra and statistics. Covers graphical representations of multivariate data matrices, multivariate distribution theory, estimation and hypothesis testing for mean vectors and variance matrices, principal components analysis, factor analysis, discrimination, cluster analysis, grouped data, ungrouped data, and latent variables. JJ

**Statistics, T, S(18).** *Statistical Process Control and Beyond.* Richard R. Clements. Robert E Krieger, 1988, xvii + 296 pp, $27.50. [ISBN: 0-89874-992-1] Good introduction to the field of Statistical Process Control (SPC). Discusses standard control charts, attribute control charts, the role of experimental design in SPC, participatory management and customer-supplier relationships. Excellent emphasis on not just using SPC methods, but using the

knowledge gained from SPC to improve the business. Continuous improvement and management involvement as keys to successful use of SPC. JJ

**Statistics, T, P, L.** *Randomized Response: Theory and Techniques.* Arijit Chaudhuri, Rahul Mukerjee. Stat.: Textbooks & Mono., V. 85. Marcel Dekker, 1988, xvi + 162 pp, $69.75. [ISBN: 0-8247-7785-9]

**Programming.** *Advanced 80386 Programming Techniques.* James L. Turley. Osborne McGraw-Hill, 1988, xv + 509 pp, $22.95 (P). [ISBN: 0-07-881342-5] A delightfully different programming book. Provides a very readable exegesis of the new features of the 80386. The assumption is that you understand the older iAPX central processing units and want to understand the value of the new features. Not so much a processor manual as an apologia. JAS

**Programming, T(13: 1).** *True BASIC by Problem Solving.* Brian D. Hahn. VCH Publishers (Suite 909, 220 E. 23rd St., NY 10010), 1988, xiv + 337 pp, $45. [ISBN: 3-527-26863-4] For the computing beginner. No mathematical prerequisites. An authentic introduction through problems: it introduces the true BASIC constructions as problems under consideration require. Many exercises, most with solutions provided. Problems from science, engineering, business, mathematics, statistics. Includes graphics, simulation, numerical methods. DFA

**Programming, S, L.** *Advanced QuickC.* Werner Feibel. Osborne McGraw-Hill, 1988, xv + 711 pp, $21.95 (P). [ISBN: 0-07-881352-2] A presentation using ANSI C plus a few PC-DOS specific extensions of standard second-level algorithms. Written for the current elementary user of C, the book avoids both "hello world" and Lampel-Zix compression. Within that broad range the coverage of topics (with coding examples) is correspondingly large: trees, lists, and the like; simulation; graph theory; statistics and random numbers; and a few QuickC specific topics. JAS

**Artificial Intelligence, S(16-18) P, L.** *Search in Artificial Intelligence.* Ed: Laveen Kanal, Vipin Kumar. Symbolic Computat. Springer-Verlag, 1988, x + 482 pp, $48. [ISBN: 0-387-96750-8] Chapters by individual authors focus on developments in search algorithms and techniques of the last decade. Special emphasis on search methods in engineering and operations research as well as artificial intelligence. Index. RJA

**Artificial Intelligence.** *Learning Issues for Intelligent Tutoring Systems.* Ed: Heinz Mandl, Alan Lesgold. Springer-Verlag, 1988, xix + 307 pp, $20 (P). [ISBN: 0-387-96616-1] Individual chapter authors. Three chapter groupings: (1) aspects of learning and instruction relevant to design of intelligent tutoring systems; (2) learning systems design; (3) conditions facilitating learning. Index. RJA

**Computer Science, P.** *Advances in Computers, V. 27.* Ed: Marshall C. Yovits. Academic Pr, 1988, xiii + 480 pp, $69. [ISBN: 0-12-012127-1] A collection of eight essentially expository essays covering such topics as military information processing, computer vision, computers in the health sciences, and computer science and information technology in the People's Republic of China. JAS

**Computer Science, P.** *Lecture Notes in Computer Science-328: VDM '88: VDM—The Way Ahead.* Ed: R. Bloomfield, L. Marshall, R. Jones. Springer-Verlag, 1988, ix + 499 pp, $37.70 (P). [ISBN: 0-387-50214-9] Proceedings of the Second Vienna Development Method (software engineering) Symposium held in Dublin, Ireland, September 11-16, 1988. JAS

**Computer Science, T, P.** *Design of Computer Data Files, Second Edition.* Owen Hanson. Computer Science Pr, 1988, xii + 419 pp, $34.95. [ISBN: 0-7167-8197-2] This edition, which still emphasizes practical concepts rather than advanced algorithm analysis or hardware and theoretical details, has added material on balanced trees, content-addressable file storage devices, and order-preserving hashing. (*First Edition*, TR, June-July 1983.) JAS

**Computer Science, S(13-14).** *Abstract Data Types: Their Specification, Representation, and Use.* Pete Thomas, Hugh Robinson, Judy Emms. Appl. Math. & Comput. Sci. Ser. Clarendon Pr, 1988, xii + 256 pp, $59.95; $26.95 (P). [ISBN: 0-19-859663-4; 0-19-859668-5] An elementary introduction to abstract data types and their use in software engineering. Two specification methods are presented: axiomatic and constructive. Examples are presented in Pascal, Modula-2, and Ada. Designed as a self-study text—solutions to nearly all of the exercises are given in an appendix. AO

**Computer Science, T(17-18: 2), S, P.** *Search Problems.* Rudolf Ahlswede, Ingo Wegener. Transl: Jean E. Wotschke. Disc. Math. & Optimiz. Wiley, 1987, xi + 284 pp, $61.95. [ISBN: 0-471-90825-8] The basic ideas, methods, and results of the field. Though introductory, much attention is given to a variety of recent applications. This volume is a translation of the 1979 Teubner Stuttgart paperback edition (TR, February 1980), but with some revisions and extensions. SS

**Computer Science, P.** *Lecture Notes in Computer Science-317: Automata, Languages and Programming.* Ed: Timo Lepistö, Arto Salomaa. Springer-Verlag, 1988, xi + 741 pp, $65.50 (P). [ISBN: 0-387-19488-6] Proceedings of the Fifteenth International Colloquium held at Tampere, Finland, July 11-15, 1988. This was a broad-based conference covering theoretical computer science with such topics as "Zeta functions of recognizable languages," and "Constructive Hopf's theorem: Or how to untangle closed planar curves." JAS

**Applications, P*, L.** *Mathematics in Industrial Problems.* Avner Friedman. Inst. for Math. & Its Applic., V. 16. Springer-Verlag, 1988, x 174 pp, $19.80. [ISBN: 0-387-96860-1] Industry offers a range of problems that mathematicians, both pure and applied, could profit from working on. A seminar at the IMA garnered a batch, from image reconstruction in oil refinery to digital syncronization. Surprisingly inexpensive—well worth the price. BC

**Applications, P.** *ICIAM '87: Proceedings of the First International Conference on Industrial and Applied Mathematics.* Ed: James McKenna, Roger Temam. SIAM, 1988, xx + 376 pp, $56.50. [ISBN: 0-89871-224-6] Proceedings of the conference held in Paris from June 29 through July 3, 1987. Contains the texts of the sixteen invited presentations as well as abstracts of the various minisymposia and a listing of the contributed papers. AO

**Applications (Biological Science), P.** *Lecture Notes in Biomathematics-75: Cell Kinetic Modelling and the Chemotherapy of Cancer.* Helmut Knolle. Springer-Verlag, 1988, viii + 157 pp, $18.90 (P). [ISBN: 0-387-50153-3] Aimed at two audiences, medical researchers and mathematicians, these notes are intended as a "guide to mathematical cell kinetics for oncologists, and also to enable mathematicians to acquire the background needed to discuss and collaborate with medical and biological researchers." In three brief chapters and a (mainly mathematical) appendix. PZ

**Applications (Biological Science), T(18: 1), L.** *The Theory of Evolution and Dynamical Systems: Mathematical Aspects of Selection.* Josef Hofbauer, Karl Sigmund. London Math. Soc. Student Texts,V. 7. Cambridge U Pr, 1988, viii + 341 pp, $65; $19.95 (P). [ISBN: 0-521-35288-6; 0-521-35838-8] Based on lectures given by the authors at the University of Vienna, this English translation of the 1984 German edition provides an introduction to the qualitative theory of ordinary differential equations, and its application to several branches of evolutionary biology: population genetics, mathematical ecology, prebiotic evolution of macromolecules, and game theoretic modeling of animal behavior. The reader is led from fundamentals to recent mathematical developments in these fields. CE

**Applications (Communication Theory), T(16-18: 2), S, L.** *Codes and Cryptography.* Dominic Welsh. Clarendon Pr, 1988, xi + 257 pp, $59.25; $32.50 (P). [ISBN: 0-19-853288-1; 0-19-853287-3] Requiring only basic courses in abstract algebra and probability, this is an introduction to the mathematics of communication theory. Covers coding, structure of natural languages, computational complexity, and much more. Exercises are really for beginners. SS

**Applications (Economics), P, L.** *A General Theory of Equilibrium Selection in Games.* John C. Harsanyi, Reinhard Selten. MIT Pr, 1988, xv + 378 pp, $32. [ISBN: 0-262-08173-3] Exposition of a new theory of bargaining that leads to a unique Nash equilibrium point, instead of to the multiplicity of non-comparable equilibria emerging from classical game theory. Following five chapters of theory are four containing applications to bargaining and multilateral trade. Robert Aumann's Foreword boldly claims that "the publication of this book constitutes a major event in game theory." LAS

**Applications (Information Theory), T(17-18: 2), S.** *Convolutional Codes: An Algebraic Approach.* Ph. Piret. MIT Pr, 1988, xiv + 339 pp, $40. [ISBN: 0-262-16110-9] An algebraic point of view is adopted to study convolutional codes in order to show its efficiency in constructing and analyzing such codes. SS

**Applications (Physical Science), P.** *Trends in Applications of Mathematics to Mechanics.* Ed: J.F. Besseling, W. Eckhaus. Springer-Verlag, 1988, ix + 361 pp, $50.20. [ISBN: 0-387-50075-8] Thirty-three papers on perturbation methods, instability, bifurcations and transition to chaos, multibody dynamics and control, and mechanics of non-classical materials. BC

**Applications (Physics), T(13: 1, 2).** *Flat and Curved Space-Times.* George F.R. Ellis, Ruth M. Williams. Clarendon Pr, 1988, x + 351 pp, $35 (P); $75. [ISBN: 0-19-851169-8; 0-19-851164-7] Amazing. General relativity explained without assuming the integral or the derivative or any physics background. Requiring only high school mathematics, the reader is gently led through special and general relativity from an algebraic and geometric perspective. Of course, a few details are omitted or relegated to appendices. All-in-all, this appears to be a reasonable introduction to Einstein's space-time accessible to any undergraduate. Many exercises but no solutions included; available upon request. Several computer programs are provided illustrating concepts. MR

**Applications (Physics), T(16-17: 1).** *Classical Equilibrium Statistical Mechanics.* Colin J. Thompson. Clarendon Pr, 1988, ix + 213 pp, $42.50. [ISBN: 0-19-851984-2] A mathematical treatment that "does not lose sight of the physics." Several problems with solutions at end of each chapter. Prerequisites: advanced calculus, probability, linear algebra, and classical mechanics. Chapters are: thermodynamics, Gibbs distribution, model systems, phase transitions, fluctuations and correlations, exactly solved models, and scaling theory and the renormalization group. SP

## Reviewers

RJA: Richard J. Allen, St. Olaf; DFA: David F. Appleyard, Carleton; SB: Steve Benson, St. Olaf; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; CE: Christopher Ennis, Carleton; SG: Steven Galovich, Carleton; GG: George Gilbert, St. Olaf; BH: Bruce Hanson, St. Olaf; RH: Rhonda Hatcher, St. Olaf; TH: Timothy Hesterberg, St. Olaf; JJ: Jason Jones, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; GN: Gail Nelson, Carleton; AO: Arnold Ostebee, St. Olaf; SP: Samuel Patterson, Carleton; MR: Matthew Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MS: Myriam Steinback, Macalester; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; LW: Lynne Walling, St. Olaf; PZ: Paul Zorn, St. Olaf.

# MAA STUDIES IN MATHEMATICS

## Studies in Combinatorics

Gian-Carlo Rota, editor
Volume #17, MAA Studies in Mathematics
272 pp., 1978, Hardbound, ISBN-0-88385-177-2
List: $26.50 MAA Member: $18.50

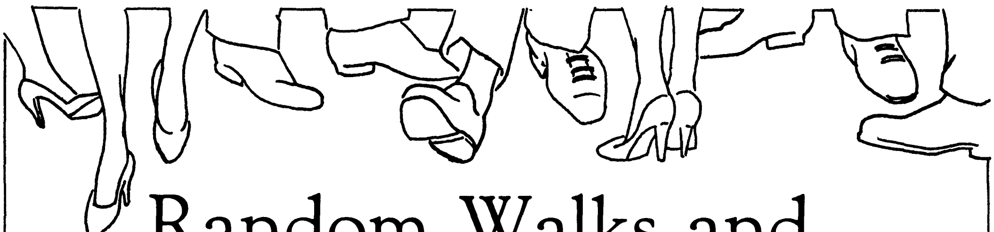
> *"This is a gem of a book. In layout and size it brings back memories of the classic* **Combinatorial Mathematics** *by H.J. Ryser. The index is very good and the editor is to be warmly congratulated on bringing out a book of enormous value and uniformly high standard. It is the sort of book which could be used with great profit as the basis of a weekly working seminar on combinatorics. I recommend it unreservedly."*
> D.J.A. Welsh in ***Bulletin of the Institute of Mathematics and its Applications***

This excellent book in the **MAA Studies in Mathematics** series contains seven papers by prominent researchers in combinatorics. Although written by different authors, the papers have the unity of chapters in a monograph. Together they present a fairly comprehensive exposition of some recent developments in combinatorics. Much of the material has application outside mathematics to computer science, operations research and network and circuit theory. Begun at the time of Euler, combinatorics has become one of the most active fields in mathematics. A significant feature of **Studies in Combinatorics** is its accessibility to any mathematician qualified to teach at the college level, as well as to a great many students. This volume is also accessible to a wide scientific public outside mathematics.

**Table of Contents**


Combinatorial Matrix Theory ............ H.J. Ryser
Proof Techniques in the Theory of Finite Sets ........ Curtis Greene and Daniel J. Kleitman
Ramsey Theory ............ R.L. Graham and B.L. Rothschild
Generating Functions ............ Richard P. Stanley
Nonconstructive Methods in Discrete Mathematics ............ Joel Spencer
Matroids and Combinatorial Geometries ............ Tom Brylawski and Douglas G. Kelly
Combinatorial Constructions ............ Marshall Hall, Jr.


Order From:
**The Mathematical Association of America**
1529 Eighteenth Street, N.W.
Washington, D.C. 20036



**Studies in the History of Mathematics,** Esther R. Phillips, Editor.
Volume #26 in MAA Studies in Mathematics
320 pp., Hardbound, ISBN-0-88385-128-8.
Catalog Number - MAS-26
List: $36.50 MAA Member: $28.00

Esther R. Phillips has brought together a collection of articles showing the sweep of recent scholarship in the history of mathematics. The material covers a wide range of current research topics: algebraic number theory, geometry, topology, logic, the relationship between mathematics and computing, partial differential equations, and algebraic geometry. This volume will show teachers and their students the historical background and the development of advanced topics in mathematics and of the major fields of contemporary research. Each article contains references for further study.


Dedekind's Invention of Ideals ................ Harold M. Edwards

Non-Euclidean Geometry and Weierstrassian Mathematics: The Background of Killing's work on Lie Algebras ..................... Thomas Hawkins

The Discovery of Non-Euclidean Geometry ........................ Jeremy Gray

L.E.J. Brouwer's Coming of Age as a Topologist .................. Dale M. Johnson

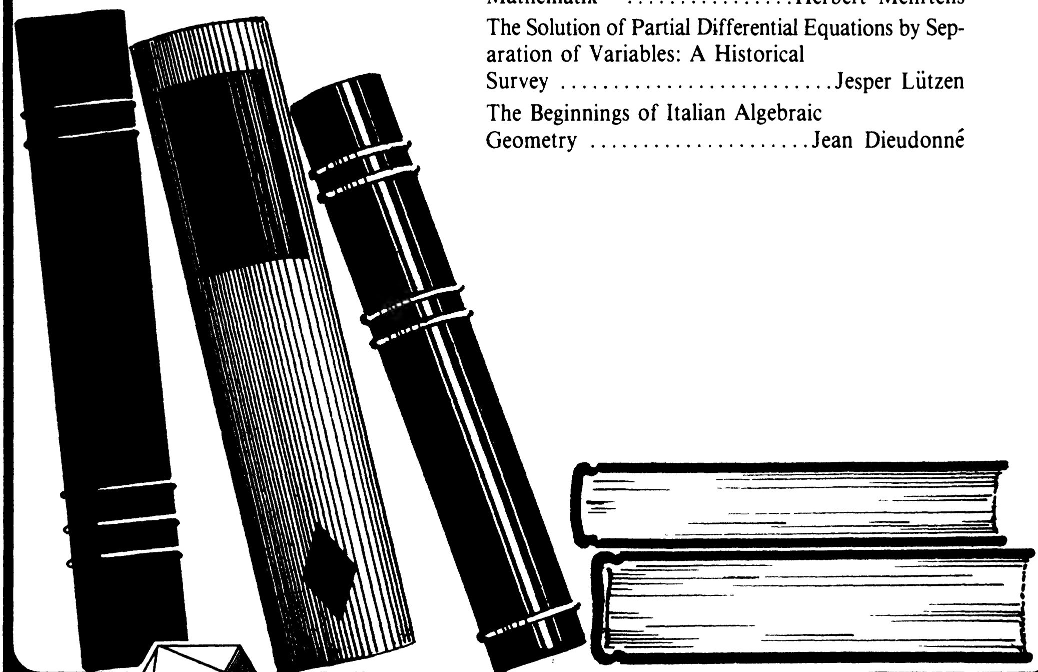
A House Divided Against Itself: The Emergence of First-Order Logic as the Basis for Mathematics ................ Gregory H. Moore

The Mathematical Reception of the Modern Computer: John von Neumann and the Institute for Advanced Study Computer ...... William Aspray

Ludwig Bieberbach and "Deutsche Mathematik" ................ Herbert Mehrtens

The Solution of Partial Differential Equations by Separation of Variables: A Historical Survey .......................... Jesper Lützen

The Beginnings of Italian Algebraic Geometry ..................... Jean Dieudonné

*Order from*
**The Mathematical Association of America**
1529 Eighteenth St. NW
Washington, D.C. 20036

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, DC 20036

THE AMERICAN
MATHEMATICAL MONTHLY

Volume 96, Number 5 May 1989

# Contents

(ISSN 0002-9890)

ARTICLES

A Motivated Proof of the Rogers-Ramanujan Identities . . . . . GEORGE E. ANDREWS AND R. J. BAXTER 401

How Not to Prove Fermat's Last Theorem . . . . . JOHN McCLEARY 410

THE EDITOR'S CORNER

Summability Theory: A Neglected Tool of Analysis . . . . . LEE A. RUBEL 421

LETTERS TO THE EDITOR . . . . . 424

UNSOLVED PROBLEMS

Conways Rats and Other Reversals . . . . . RICHARD K. GUY 425

NOTES

The Problem of Câlissons . . . . . GUY DAVID AND CARLOS TOMEI 429

Arakelian's Approximation Theorem . . . . . JEAN-PIERRE ROSAY AND WALTER RUDIN 432

The Norm of a Linear Function . . . . . I. J. MADDOX 434

An Elementary Treatment of the Radon-Nikodym Derivative . . . . . RICHARD C. BRADLEY 437

THE TEACHING OF MATHEMATICS

On the Proof of the Radon-Nikodym Theorem . . . . . ALBERT WILANSKY 441

The Butterfly Curve . . . . . TEMPLE H. FAY 442

Nonstandard Continuity and Uniform Convergence . . . . . CHRISTOPHER L. THOMPSON 443

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions . . . . . 445

Advanced Problems and Solutions . . . . . 454

REVIEWS

Algorithms in Combinatorial Geometry. By Herbert Edelsbrunner . . . . . JACOB E. GOODMAN 457

Elementary Number Theory. By Charles Vanden Eynden . . . . EMIL GROSSWALD 460

Combinatorics of Finite Sets. By Ian Anderson . . . . . DANIEL J. KLEITMAN 463

For All Practical Purposes: An Introduction to Contemporary Mathematics. By COMAP . . . . . 465

TELEGRAPHIC REVIEWS . . . . . 468

# NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to ANITA E. SOLOW, Department of Mathematics, Grinnell College, Grinnell, IA 50112; in geometry or topology to DENNIS DETURCK, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems (three copies) and solutions (two copies), both elementary and advanced to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, University of Calgary, Alberta, Canada T2N 1N4



Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.


The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $26 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $29. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $90 per year.





Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Assocation of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

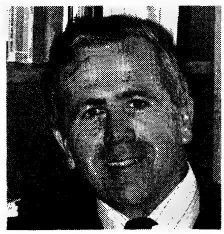
# A Motivated Proof of the Rogers-Ramanujan Identities

GEORGE E. ANDREWS,[1] *Pennsylvania State University, University Park, PA 16802*
R. J. BAXTER, *Australian National University, Canberra, Australia*

GEORGE ANDREWS received his Ph.D. at the University of Pennsylvania in 1964. He was the late Professor Hans Rademacher's last student and is currently Evan Pugh Professor of Mathematics at the Pennsylvania State University.

He has spent leaves of absence at MIT, University of Wisconsin, University of New South Wales, the Australian National University and the University of Strasbourg. He has published extensively on the theory of partitions and related areas of mathematics. His most recent book is *q-Series*, a monograph derived from his Regional Conference Lectures at Arizona State University in 1985.

RODNEY BAXTER was an undergraduate at Cambridge, England, and went to Australia to do a Ph.D. in 1961. Afterwards he worked briefly for an oil company. He taught in the Math Department at M.I.T. from 1968 to 1970, and since then has been at the Australian National University. His speciality is exact solutions in statistical mechanics: he received the 1987 Heineman Prize for Mathematical Physics from the American Physical Society.



**1. Introduction.** The Rogers-Ramanujan identities are most simply stated as assertions about partitions [**1**, p. 109].

*The First Rogers-Ramanujan Identity.* The partitions of $n$ with difference between parts at least 2 are equinumerous with the partitions of $n$ into parts of the forms $5m + 1$ and $5m + 4$.

*The Second Rogers-Ramanujan Identity.* The partitions of $n$ into parts $> 1$ with difference between parts at least 2 are equinumerous with the partitions of $n$ into parts of the forms $5m + 2$ and $5m + 3$.

As an example of the first identity, when $n = 9$ there are five partitions with difference at least 2 between parts: $9, 8 + 1, 7 + 2, 6 + 3, 5 + 3 + 1$, and there are five partitions whose parts are of the forms $5m + 1$ and $5m + 4$: $9, 6 + 1 + 1 + 1, 4 + 4 + 1, 4 + 1 + 1 + 1 + 1 + 1, 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1$. As an example of the second identity, when $n = 12$ there are six partitions into parts $> 1$ with difference at least 2 between parts: $12, 10 + 2, 9 + 3, 8 + 4, 7 + 5, 6 + 4 + 2$, and there are six partitions whose parts are of the forms $5m + 2$ and $5m + 3$: $12, 8 + 2 + 2, 7 + 3 + 2, 3 + 3 + 3 + 3, 3 + 3 + 2 + 2 + 2, 2 + 2 + 2 + 2 + 2 + 2$.

The surprising history of these theorems was given by G. H. Hardy [**8**, p. 91], and their place in recent work was described in [**2**]. They were first proved as analytic identities in 1894 by L. J. Rogers [**11**]; however his work was soon forgotten. In 1913 Ramanujan, having discovered them empirically, sent them to Hardy as conjectures. In 1916, MacMahon [**9**, Ch. III] discussed them at length as conjectures and presented them in the form given above. After the publication of MacMahon's

[1] Partially supported by National Science Foundation Grant DMS-8503324.

book, Ramanujan ran across Rogers' 1894 paper. Subsequent correspondence between them resulted in [10]. I. Schur [11] independently discovered these identities in 1917.

It is an easy matter to show that

$$\begin{aligned} G_1(q) &\equiv \prod_{m=0}^{\infty} \frac{1}{(1-q^{5m+1})(1-q^{5m+4})} \\ &= 1 + q + q^2 + q^3 + 2q^4 + 2q^5 + 3q^6 + 3q^7 + 4q^8 + 5q^9 + \cdots \end{aligned} \tag{1.1}$$

is the generating function for partitions of $n$ into parts of the forms $5m+1$ and $5m+4$ [1; Ch. 1], and

$$\begin{aligned} G_2(q) &= \prod_{m=0}^{\infty} \frac{1}{(1-q^{5m+2})(1-q^{5m+3})} \\ &= 1 + q^2 + q^3 + q^4 + q^5 + 2q^6 + 2q^7 + 3q^8 + 3q^9 + \cdots \end{aligned} \tag{1.2}$$

is the generating function for partitions of $n$ into parts of the forms $5m+2$ and $5m+3$. At the 1987 A.M.S. Institute on Theta Functions, Leon Ehrenpreis asked if one could prove that

$$\begin{aligned} G_1(q) - G_2(q) &= \prod_{n=0}^{\infty} \frac{1}{(1-q^{5n+1})(1-q^{5n+4})} - \prod_{n=0}^{\infty} \frac{1}{(1-q^{5n+2})(1-q^{5n+3})} \\ &= q(1 + q^3 + q^4 + q^5 + q^6 + q^7 + 2q^8 + 2q^9 + \cdots) \\ &\equiv qG_3(q) \end{aligned} \tag{1.3}$$

has nonnegative coefficients without resorting to the Rogers-Ramanujan identities. In answering this question we are led naturally to the Rogers-Ramanujan identities themselves. Indeed the journey we take is the way in which one of us (RJB) independently discovered and proved the Rogers-Ramanujan identities for application in statistical mechanics [3], [4], [5].

**2. Empirical Derivation and Ehrenpreis's Problem.** To be specific, we are given $G_1(q)$ and $G_2(q)$ as the above infinite products. We observe with Ehrenpreis that $G_3(q)$ defined in (1.3) has nonnegative coefficients. Furthermore we observe that

$$\begin{aligned} G_2(q) - G_3(q) &= q^2(1 + q^4 + q^5 + q^6 + q^7 + q^8 + q^9 + 2q^{10} + \cdots) \\ &\equiv q^2 G_4(q), \end{aligned} \tag{2.1}$$

and again $G_4(q)$ has nonnegative coefficients. Intrigued we try the same thing again, and now we find

$$G_3(q) - G_4(q) = q^3(1 + q^5 + q^6 + q^7 + q^8 + q^9 + q^{10} + \cdots) \equiv q^3 G_5(q), \tag{2.2}$$

and $G_5(q)$ has nonnegative coefficients as well. Indeed it appears that we might be interested in the infinite family of functions $G_i(q)$ given by

$$G_i(q) = \frac{G_{i-2}(q) - G_{i-1}(q)}{q^{i-2}}, \qquad i \geqslant 3. \tag{2.3}$$

Empirically we find that each $G_i(q)$ has no negative powers of $q$, and the following table provides the first few coefficients of $G_i(q)$ with $1 \leqq i \leqq 8$

| $n$ | $G_1(q)$ | $G_2(q)$ | $G_3(q)$ | $G_4(q)$ | $G_5(q)$ | $G_6(q)$ | $G_7(q)$ | $G_8(q)$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 4 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| 5 | 2 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| 6 | 3 | 2 | 1 | 1 | 1 | 1 | 0 | 0 |
| 7 | 3 | 2 | 1 | 1 | 1 | 1 | 1 | 0 |
| 8 | 4 | 3 | 2 | 1 | 1 | 1 | 1 | 1 |
| 9 | 5 | 3 | 2 | 1 | 1 | 1 | 1 | 1 |
| 10 | 6 | 4 | 3 | 2 | 1 | 1 | 1 | 1 |
| 11 | 7 | 4 | 3 | 2 | 1 | 1 | 1 | 1 |
| 12 | 9 | 6 | 4 | 3 | 2 | 1 | 1 | 1 |
| 13 | 10 | 6 | 4 | 3 | 2 | 1 | 1 | 1 |
| 14 | 12 | 8 | 5 | 4 | 3 | 2 | 1 | 1 |
| 15 | 14 | 9 | 6 | 4 | 3 | 2 | 1 | 1 |
| 16 | 17 | 11 | 7 | 5 | 4 | 3 | 2 | 1 |
| 17 | 19 | 12 | 8 | 5 | 4 | 3 | 2 | 1 |
| 18 | 23 | 15 | 10 | 7 | 5 | 4 | 3 | 1 |
| 19 | 26 | 16 | 11 | 7 | 5 | 4 | 3 | 2 |
| 20 | 31 | 20 | 13 | 9 | 6 | 5 | 4 | 2 |

The above table (and a much more extensive one) clearly suggests that the top row consists entirely of ones and if the top row is deleted the remaining array is an infinite lower triangular matrix. More explicitly, we may propose the

**Empirical Hypothesis.**

$$G_i(q) = 1 + \sum_{n=i}^{\infty} g_{i,n} q^n. \tag{2.4}$$

It is now only a short distance from (2.4) to a positive answer to Ehrenpreis's question. Namely we wish to show that all the coefficients of $G_3(q)$ are nonnegative. To do this we write $G_3(q)$ as a linear combination of $G_i(q)$ and $G_{i+1}(q)$ for each $i \geqq 3$. We can easily determine this combination by repeatedly applying (2.3) rewritten as

$$G_{i-2}(q) = G_{i-1}(q) + q^{i-2}G_i(q), \qquad i \geqq 3. \tag{2.5}$$

Thus

$$\begin{aligned} G_3(q) &= G_4(q) + q^3 G_5(q) \quad \text{(by (2.5) with } i = 5) \\ &= \left(G_5(q) + q^4 G_6(q)\right) + q^3 G_5(q) \quad \text{(by (2.5) with } i = 6) \\ &= (1 + q^3)G_5(q) + q^4 G_6(q) \\ &= (1 + q^3 + q^4)G_6(q) + q^5(1 + q^3)G_7(q) \quad \text{(by (2.5) with } i = 7). \end{aligned}$$

Proceeding in this manner we see that

$$G_3(q) = A_i(q)G_i(q) + B_i(q)G_{i+1}(q), \tag{2.6}$$

where (by (2.5))

$$A_{i+1}(q) = A_i(q) + B_i(q) \tag{2.7}$$

and

$$B_{i+1}(q) = q^i A_i(q) \tag{2.8}$$

with $A_3(q) = 1$, $B_3(q) = 0$.

Eliminating $B_i(q)$ from (2.7) and (2.8) we find $A_3(q) = 1$, $A_4(q) = 1$, and for $i \geqq 4$,

$$A_{i+1}(q) = A_i(q) + q^{i-1}A_{i-1}(q). \tag{2.9}$$

This recurrence and the initial conditions easily establish that for $|q| < 1$, $A_i(q)$ converges as $i \to \infty$ to a power series with nonnegative coefficients. Therefore if (2.4) is true, then by (2.6)

$$\begin{aligned} G_3(q) &= \lim_{i\to\infty} \left(A_i(q)G_i(q) + q^{i-1}A_{i-1}(q)G_{i+1}(q)\right) \\ &= \left(\lim_{i\to\infty} A_i(q)\right)\left(\lim_{i\to\infty} G_i(q)\right) + 0\left(\lim_{i\to\infty} A_{i-1}(q)\right)\left(\lim_{i\to\infty} G_{i+1}(q)\right) \\ &= A_\infty(q)\cdot 1 + 0\cdot A_\infty(q)\cdot 1 \quad \text{(by (2.4))} \\ &= A_\infty(q). \end{aligned} \tag{2.10}$$

Hence by the remarks following (2.9) we see that $G_3(q) = A_\infty(q)$ does indeed have nonnegative coefficients assuming (2.4) is true.

**3. Proof of the Empirical Hypothesis.** The answer to Ehrenpreis's question given in Section 2 relies on the unproved observation (2.4). Now we would like to discover representations of the $G_i(q)$ which would allow us to prove (2.4). Our starting point is the famous Triple Product Identity of Jacobi:

$$\sum_{\lambda=-\infty}^{\infty} (-1)^\lambda q^{\lambda^2} z^\lambda = \prod_{n=0}^{\infty} \left\{(1-q^{2n+2})(1-zq^{2n+1})(1-z^{-1}q^{2n+1})\right\}. \tag{3.1}$$

There are simple proofs of this identity given in [**1**, p. 21], [**2**, p. 115]. Consequently with $q$ replaced by $q^{5/2}$ and $z$ by $q^{1/2}$, we obtain

$$\begin{aligned} G_1(q) &= \frac{1 + \sum_{\lambda=1}^{\infty} (-1)^\lambda q^{\lambda(5\lambda-1)/2}(1+q^\lambda)}{\prod_{n=1}^{\infty}(1-q^n)} \\ &= \frac{1 - q^2 - q^3 + q^9 + q^{11} - q^{21} - q^{24} + q^{38} + q^{42} - \cdots}{\prod_{n=1}^{\infty}(1-q^n)}, \end{aligned} \tag{3.2}$$

and with $q$ replaced by $q^{5/2}$ and $z$ replaced by $q^{3/2}$, we obtain

$$\begin{aligned} G_2(q) &= \frac{\sum_{\lambda=0}^{\infty} (-1)^\lambda q^{\lambda(5\lambda+3)/2}(1-q^{2\lambda+1})}{\prod_{n=1}^{\infty}(1-q^n)} \\ &= \frac{1 - q - q^4 + q^7 + q^{13} - q^{18} - q^{27} + q^{34} + q^{46} - \cdots}{\prod_{n=1}^{\infty}(1-q^n)}. \end{aligned} \tag{3.3}$$

Let us use (3.2) and (3.3) to find $G_3(q)$: By (2.5)

$$\begin{aligned}
qG_3(q) &= G_1(q) - G_2(q)\\
&= \Big\{\big(1 - q^2 - q^3 + q^9 + q^{11} - q^{21} - q^{24} + q^{38} + q^{42} - \cdots\big)\\
&\quad -\big(1 - q - q^4 + q^7 + q^{13} - q^{18} - q^{27} + q^{34} + q^{46} - \cdots\big)\Big\}\Big/\prod_{n=1}^{\infty}(1-q^n)\\
&= \frac{\begin{array}{l}(q - q^2 - q^3 + q^4 - q^7 + q^9 + q^{11} - q^{13}\\ \quad + q^{18} - q^{21} - q^{24} + q^{27} - q^{34} + q^{38} + q^{42} - q^{46} \cdots)\end{array}}{\prod_{n=1}^{\infty}(1-q^n)}\\
&= \frac{\begin{array}{l}q\big((1-q)(1-q^2) - q^6(1-q^2)(1-q^4)\\ \quad + q^{17}(1-q^3)(1-q^6) - q^{33}(1-q^4)(1-q^8)\cdots\big)\end{array}}{\prod_{n=1}^{\infty}(1-q^n)}.
\end{aligned} \tag{3.4}$$

Thus we see that

$$G_3(q) = \frac{\sum_{\lambda=0}^{\infty}(1-q^{\lambda+1})(1-q^{2\lambda+2})(-1)^{\lambda}q^{6\lambda+5\binom{\lambda}{2}}}{\prod_{n=1}^{\infty}(1-q^n)}. \tag{3.5}$$

Since (3.5) is so nice, let us use (3.3) and (3.4) to find $G_4(q)$. By (2.5)

$$\begin{aligned}
q^2G_4(q) &= G_2(q) - G_3(q)\\
&= \Big\{\big(1 - q - q^4 + q^7 + q^{13} - q^{18} - q^{27} + q^{34} + q^{46} - \cdots\big)\\
&\quad -\big(1 - q - q^2 + q^3 - q^6 + q^8 + q^{10} - q^{12} + q^{17} - q^{20} - q^{23} + q^{26}\\
&\quad - q^{33} + q^{37} + q^{41} - q^{45} + \cdots\big)\Big\}\Big/\prod_{n=1}^{\infty}(1-q^n)\\
&= q^2\big(1 - q - q^2 + q^4 + q^5 - q^6 - q^8 + q^{10} + q^{11} - q^{15} - q^{16} + q^{18}\\
&\quad + q^{21} - q^{24} - q^{25} + q^{31} + q^{32} - q^{35} - q^{39} + q^{43} + q^{44} \cdots\big)\Big/\sum_{n=1}^{\infty}(1-q^n)\\
&= q^2\big((1-q)(1-q^2)(1-q^3) - q^8(1-q^2)(1-q^3)(1-q^5)\\
&\quad + q^{21}(1-q^3)(1-q^4)(1-q^7) - q^{39}\cdots\big)\Big/\prod_{n=1}^{\infty}(1-q^n).
\end{aligned} \tag{3.6}$$

Therefore, we see that

$$G_4(q) = \frac{\sum_{\lambda=0}^{\infty}(1-q^{\lambda+1})(1-q^{\lambda+2})(1-q^{2\lambda+3})(-1)^{\lambda}q^{8\lambda+5\binom{\lambda}{2}}}{\prod_{n=1}^{\infty}(1-q^n)}. \tag{3.7}$$

From (3.3), (3.5) and (3.7) we may conjecture that

$$G_i(q) = \frac{\sum_{\lambda=0}^{\infty}(1-q^{\lambda+1})(1-q^{\lambda+2})\cdots(1-q^{\lambda+i-2})(-1)^{\lambda}q^{2i\lambda+5\binom{\lambda}{2}}(1-q^{2\lambda+i-1})}{\prod_{n=1}^{\infty}(1-q^n)}. \tag{3.8}$$

Furthermore it is not complicated to prove (3.8) in general if we carefully examine the intermediate steps in the derivation of (3.6) to see how the various terms fit together. Namely, assuming (3.8) for $i$ and $i+1$, we see that

$$\begin{aligned}
&\prod_{n=1}^{\infty}(1-q^n)(G_i(q)-G_{i+1}(q))\\
&= (1-q)(1-q^2)\cdots(1-q^{i-1}) + \sum_{\lambda=1}^{\infty}(1-q^{\lambda+1})(1-q^{\lambda+2})\\
&\quad\times\cdots(1-q^{\lambda+i-2})(-1)^{\lambda}q^{2i\lambda+5\binom{\lambda}{2}}(1-q^{2\lambda+i-1})\\
&\quad-\left\{\sum_{\lambda=0}^{\infty}(1-q^{\lambda+1})\cdots(1-q^{\lambda+i-1})(-1)^{\lambda}q^{2(i+1)\lambda+5\binom{\lambda}{2}}\right.\\
&\qquad\left.-\sum_{\lambda=0}^{\infty}(1-q^{\lambda+1})\cdots(1-q^{\lambda+i-1})(-1)^{\lambda}q^{2(i+1)\lambda+5\binom{\lambda}{2}+2\lambda+i}\right\}\\
&= -\sum_{\lambda=0}^{\infty}(1-q^{\lambda+2})(1-q^{\lambda+3})\\
&\quad\times\cdots(1-q^{\lambda+i-1})(-1)^{\lambda}q^{2i(\lambda+1)+5\binom{\lambda+1}{2}}(1-q^{2\lambda+i+1})\\
&\quad+\sum_{\lambda=0}^{\infty}(1-q^{\lambda+2})(1-q^{\lambda+3})\cdots(1-q^{\lambda+i})(-1)^{\lambda}q^{2(i+1)(\lambda+1)+5\binom{\lambda+1}{2}}\\
&\quad+\sum_{\lambda=0}^{\infty}(1-q^{\lambda+1})\cdots(1-q^{\lambda+i-1})(-1)^{\lambda}q^{2(i+1)\lambda+5\binom{\lambda}{2}+2\lambda+i}\\
&= q^i\sum_{\lambda=0}^{\infty}(1-q^{\lambda+2})\cdots(1-q^{\lambda+i-1})(-1)^{\lambda}q^{2(i+1)\lambda+5\binom{\lambda}{2}+2\lambda}\\
&\quad\left(-q^{i+\lambda}(1-q^{2\lambda+i+1})+q^{i+2+3\lambda}(1-q^{\lambda+i})+(1-q^{\lambda+1})\right)\\
&= q^i\sum_{\lambda=0}^{\infty}(1-q^{\lambda+2})\cdots(1-q^{\lambda+i-1})\\
&\quad\times(-1)^{\lambda}q^{2(i+2)\lambda+5\binom{\lambda}{2}}\left((1-q^{\lambda+1})(1-q^{\lambda+i})(1-q^{2\lambda+i+1})\right)\\
&= q^i\sum_{\lambda=0}^{\infty}(1-q^{\lambda+1})(1+q^{\lambda+2})\cdots(1-q^{\lambda+i})(-1)^{\lambda}q^{2(i+2)\lambda+5\binom{\lambda}{2}}\\
&\quad\times(1-q^{2\lambda+i+1})\left(=q^iG_{i+2}(q)\prod_{n=1}^{\infty}(1-q^n)\right).
\end{aligned} \tag{3.9}$$

Thus we see that the right-hand side of (3.8) is indeed equal to $G_i(q)$ for $i = 2$, 3, and 4 (by (3.3), (3.5) and (3.7)) and (3.9) shows that the right-hand side fulfills the second-order recurrence (2.5). Hence (3.8) is valid for all $i \geqq 2$.

The Empirical Hypothesis is an immediate consequence of (3.8). Namely, by (3.8), we see that

$$G_i(q) = \frac{1}{\prod_{n=1}^{\infty}(1-q^{n+i-1})} + \frac{\sum_{\lambda=1}^{\infty}(1-q^{\lambda+1})\cdots(1-q^{\lambda+i-2})(-1)^{\lambda}q^{2i\lambda+5\binom{\lambda}{2}}(1-q^{2\lambda+i-1})}{\prod_{n=1}^{\infty}(1-q^n)}$$
$$= 1 + q^i\gamma_i(q), \tag{3.10}$$

where $\gamma_i(q)$ is analytic in $q$ around $q = 0$. I.e. the Empirical Hypothesis (2.4) is valid.

**4. The Rogers-Ramanujan Identities.** If, instead of considering $G_3(q)$, we consider $G_1(q)$, we obtain exactly the same conclusions the only variation being that the $A_i(q)$ satisfy $A_2(q) = 1$, $A_3(q) = 1 + q$ and for $i > 3$, (2.9) is still valid. Iterating (2.9), we find

$$A_4(q) = 1 + q + q^2$$
$$A_5(q) = 1 + q + q^2 + q^3 + q^{3+1}$$
$$A_6(q) = 1 + q + q^2 + q^3 + q^{3+1} + q^4 + q^{4+1} + q^{4+2}.$$

These examples suggest that $A_{i+1}(q)$ is the generating function for partitions with difference between parts at least 2 and largest part $< i$. Indeed this is easily proved by mathematical induction applied to (2.9) since the $A_i(q)$ yields the desired partitions with largest part $< i - 1$ and the $q^{i-1}A_{i-1}(q)$ yields the remaining desired partitions having largest part equal to $i - 1$.

Just as in Section 2, we see that

$$G_1(q) = A_\infty(q). \tag{4.1}$$

Now (4.1) is merely the restatement of the first Rogers-Ramanujan identity in terms of generating functions; namely $G_1(q)$ is the generating function for partitions of $n$ with parts of the forms $5m + 1$ and $5m + 4$ while $A_\infty(q)$ is the generating function for partitions with parts differing by at least 2.

The derivation of the second Rogers-Ramanujan identity follows in exactly the same manner from the Empirical Hypothesis. In particular we find that

$$G_2(q) = \alpha_i(q)G_i(q) + \beta_i(q)G_{i+1}(q), \tag{4.2}$$

where

$$\alpha_{i+1}(q) = \alpha_i(q) + \beta_i(q) \tag{4.3}$$

and

$$\beta_{i+1}(q) = q^i\alpha_i(q) \tag{4.4}$$

with $\alpha_3(q) = 1$ and $\beta_3(q) = q^2$.

Proceeding as before we find that the Empirical Hypothesis implies that

$$G_2(q) = \alpha_\infty(q) \tag{4.5}$$

and $\alpha_\infty(q)$ is the generating function for partitions with parts $> 1$ and difference at least 2 between parts.

**5. Conclusion and Comments.** It should be pointed out that the work in Section 3 is essentially the way in which one of us (RJB) independently discovered the Rogers-Ramanujan identities in connection with the original solution of the Hard Hexagon Model [**3**], [**5**, ch. 14] (cf. [**2**, Chs. 1 and 8]). The Hard Hexagon Model leads directly to (2.5) together with the series expansion

$$G_i(q) = \sum_{n \geq 0} \frac{q^{n^2+(i-1)n}}{(1-q)(1-q^2)\cdots(1-q^n)}. \tag{5.1}$$

An application of Euler's algorithm [**2**, Ch. 10, Th. 10.3] for the determination of the $a_n$ in

$$1 + \sum_{n=1}^{\infty} b_n q^n = \prod_{n=1}^{\infty} (1-q^n)^{-a_n} \tag{5.2}$$

led to the Rogers-Ramanujan identities as conjectures. The remaining empirical work and derivation in Section 3 followed. Finally (5.1) and (3.8) are identified by observing that each is an analytic function $f(z)$ in $z = q^{i-1}$ satisfying the boundary condition $f(0) = 1$ (i.e., $G_\infty(q) = 1$) and

$$f(z) = f(zq) + zqf(zq^2). \tag{5.3}$$

Even though we have only treated the sequence $z = q^{i-1}$ this is adequate since it is an infinite sequence converging to 0 for $|q| < 1$.

This approach yields the Rogers-Ramanujan identities in their more familiar analytic form [**1**, p. 104]:

$$G_1(q) = \sum_{n=0}^{\infty} \frac{q^{n^2}}{(1-q)(1-q^2)\cdots(1-q^n)} = \prod_{n=0}^{\infty} \frac{1}{(1-q^{5n+1})(1-q^{5n+4})}, \tag{5.4}$$

and

$$G_2(q) = \sum_{n=0}^{\infty} \frac{q^{n^2+n}}{(1-q)(1-q^2)\cdots(1-q^n)} = \prod_{n=0}^{\infty} \frac{1}{(1-q^{5n+2})(1-q^{5n+3})}. \tag{5.5}$$

We should point out also that the crucial recurrence for (3.8) derived in (3.9) is essentially equivalent to the proofs given by Rogers and Ramanujan in [**10**]. The relationship between our treatment and that found in [**10**] parallels the relationship between I. J. Good's formula [7] for the infinite product $\prod_{n \geq 1}(1-q^n)$ and Sylvester's identity [**1**, p. 140, Th. 9.2] (cf. [**6**]).

It may be objected that we have presented a somewhat stilted motivation. Indeed if (4.1) were not in the back of our minds, we would never have thought to construct Table 1. This may well be; however, in this age of the computer, the construction of

tables of successive differences is standard procedure when studying sequences with little given information. Furthermore, we emphasize that Table 1 is purely inherent in (1.1) and (1.2) and may well suggest a reasonable approach to the study of other pairs of infinite products where one may suspect that interesting facts in additive number theory are somewhere around.

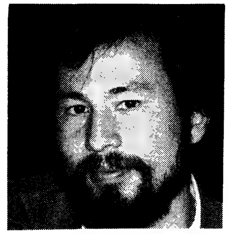
REFERENCES

1. G. E. Andrews, The Theory of Partitions, Encyclopedia of Mathematics and Its Applications, G.-C. Rota ed., Vol. 2, Addison-Wesley, Reading, 1976 (Reprinted, Cambridge University Press, London and New York, 1984).
2. ______, *q*-Series: Their Development and Application in Analysis, Number Theory, Combinatorics, Physics, and Computer Algebra, C.B.M.S. Regional Conference Series in Mathematics, No. 66, Amer. Math. Soc., Providence, 1986.
3. R. J. Baxter, Hard hexagons: exact solution, *J. Phys. A*, 13 (1980) L61-L70.
4. ______, Rogers-Ramanujan identities in the hard hexagon model, *J. Stat. Phys.*, 26 (1981) 427–452.
5. ______, Exactly Solved Models in Statistical Mechanics, Academic Press, London and New York, 1982.
6. D. Benson, W. Feit and R. Howe, Finite linear groups, the Commodore 64, Euler and Sylvester, *Amer. Math. Monthly*, 93 (1986) 717–719.
7. I. J. Good, A generalization of the Bernoulli-Euler partition formula, *Scripta Math.*, 28 (1968) 319–320.
8. G. H. Hardy, Ramanujan, Cambridge University Press, 1940 (Reprinted: Chelsea, New York, 1959).
9. P. A. MacMahon, Combinatory Analysis, Vol. 2, Cambridge University Press, London, 1916 (Reprinted, Chelsea, New York, 1960).
10. S. Ramanujan and L. J. Rogers, Proof of certain identities in combinatory analysis, *Proc. Cambridge Phil. Soc.*, 19 (1919) 211–214.
11. L. J. Rogers, Second memoir on the expansion of certain infinite products, *Proc. London Math. Soc.*, 25 (1984) 318–343.
12. I. Schur, Ein Beitrag zur additiven Zahlentheorie und zur Theorie der Kettenbrüche, S.-B. Preuss. Akad. Wiss. Phys.-Math. Kl., pp. 302–321. (Reprinted, in I. Schur, Gesammelte Abhandlungen, Vol. 2, Springer, Berlin, 1973, pp. 117–136).

# How Not to Prove Fermat's Last Theorem

JOHN MCCLEARY, *Vassar College, Poughkeepsie, NY 12601*

JOHN MCCLEARY received his Ph.D. in mathematics from Temple University in 1979. He is presently an associate professor at Vassar College. He has been a visitor to the Mathematics Institute in Göttingen and the University of Sydney. His principal area of research is algebraic topology, but number theory is a calling he cannot resist. He is the author of *User's Guide to Spectral Sequences*.



*To the memory of Emil Grosswald*

**1. Introduction.** This article is concerned with one of the oldest unsolved problems in mathematics:

FERMAT'S LAST THEOREM. *If* $n \geqslant 3$ *and* $x$, $y$, $z$ *are integers satisfying the equation* $x^n + y^n = z^n$, *then* $xyz = 0$.

The history of the search for a proof of this theorem is filled with many partial successes and many complete failures. In this article, I report on my own attempt and failure because I feel that the method is of considerable interest in spite of its lack of success.

The central idea revolves around the question of how well one can approximate a given real number, $\alpha$, with rational numbers. The goodness of the approximation can be used to determine whether or not $\alpha$ is a rational. This method has been used recently for a number of interesting problems in number theory, and with great success (see [**1**] for a survey of the work of Apéry, Alan Baker, the Chudnovskys and Beukers).

In applying this idea to Fermat's Last Theorem, we reduce the problem to the study of the rationality of $n$th roots of certain rational numbers. A tried and true method of approximation, Newton's method, leads to a sequence of rational approximants that we try to show converge too quickly to be giving a rational number. It is at this final stage that the approach fails.

In section 2 we reformulate Fermat's Last Theorem suitably. In section 3 criteria are presented for recognizing when a real number is irrational; one criterion treats the rate of approximation by rational numbers, and the other concerns the continued fraction representation of the number. In section 4 Newton's Method is introduced and the Main Idea stated, followed by an example where the Main Idea is shown to work by playing the Newton approximation off against the continued fractions approximation. In section 5 we review the facts about the convergence of Newton's Method as it applies to the problem at hand. This leads to an inequality that is crucial to the success or failure of the Main Idea. In section 6 the explicit form of the Newton approximants is considered along with the growth of their denominators. Another inequality results that, together with the previous inequality, leads us to the main theorem of the paper, and the extent to which the Main Idea can be made to work.

**2. A restatement of the problem.** Since Fermat's Last theorem is one of the most famous unsolved problems in mathematics, there is much written about it. For the

story of its appearance and development see the excellent books of Edwards [3] and Ribenboim [7]. According to Edwards [3], the first known reference to this conjecture is in a letter of Fermat to Mersenne of June, 1638, that is, 350 years ago. Fermat did leave indications of a proper proof in the case of $n = 4$, Euler gave a complete proof for the case $n = 3$ in 1753, and Dirichlet and Legendre gave a proof for the case $n = 5$ in 1825.

In order to motivate the approach in this paper, we need a restatement of the problem that will be amenable to methods outside of elementary number theory. To begin, notice that a solution to the Diophantine equation, $x^n + y^n = z^n$ with $x$, $y$, $z$ integers, by dividing both sides of the equation we get a point on the curve $X^n + Y^n = 1$ in the plane with the property that $X$ and $Y$ are rationals. Call such a point a rational point and let $C_n$ be the curve in $\mathbf{R}^2$ defined by

$$C_n = \{(X, Y) \in \mathbf{R}^2 | X^n + Y^n = 1\}.$$

From the definition of this curve, Fermat's Last Theorem can now be restated:

*For all $n \geqslant 3$, the only rational points on the curve $C_n$ are among the points $(0, \pm 1)$ or $(\pm 1, 0)$.*

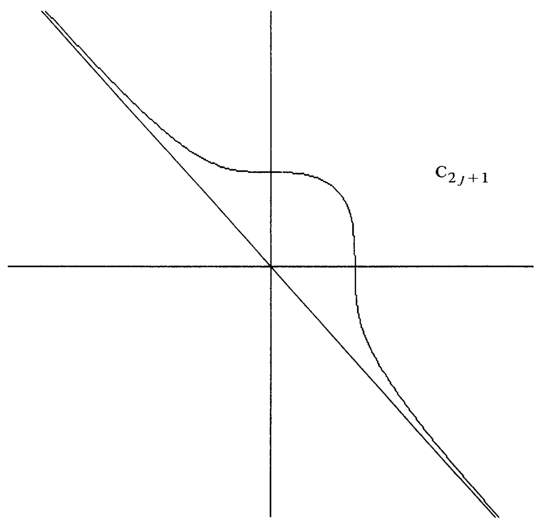



FIG. 1

Restricting to the solutions that consist of positive integers, we need to consider only rational points with the $X$-coordinate lying between 0 and 1. This leads to the following:

PROPOSITION 1. *Fermat's Last Theorem holds if and only if, for all $n \geqslant 3$ and all rationals $p/q$ with $0 < p/q < 1$, $\sqrt[n]{1 - (p/q)^n}$ is an irrational number.*

In this formulation Fermat's Last Theorem is about the irrationality of the $n$th roots of particular rational expressions. This leads to an approach to proving the conjecture to be investigated below.

**3. Recognizing irrationals.** If $r$ is a real number, how can one determine whether it is irrational or not? This problem is the subject of an entire branch of number

theory—Diophantine approximation. A fundamental insight in this area is that rationals cannot approximate rational numbers very well.

PROPOSITION 2. *Suppose $\alpha = m/n$, a rational number, in lowest terms. If $\varepsilon > 0$, then there are only finitely many rationals, $p/q$, such that*

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^{1+\varepsilon}}.$$

*Proof.* Consider

$$\left|\frac{m}{n} - \frac{p}{q}\right| = \frac{|mq - pn|}{|n|} \cdot \frac{1}{|q|}.$$

For $p/q \neq m/n$ we have $|mq - pn| \geqslant 1$ and so

$$\left|\frac{m}{n} - \frac{p}{q}\right| \geqslant \frac{1}{|n|\,|q|} > \frac{1}{2|n|\,|q|}.$$

Only finitely many $q$ satisfy $2|n| > |q|^{\varepsilon}$, and for each such $q$, there are only finitely many $p$ with

$$\frac{|mq - pn|}{|n|} < \frac{1}{|q|^{\varepsilon}}.$$

This proves the proposition.

COROLLARY 3. *If $\alpha$ is a real number, and there exists an $\varepsilon > 0$ and a sequence of distinct rationals, $\{p_n/q_n\}$, converging to $\alpha$ and satisfying*

$$\left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{q_n^{1+\varepsilon}},$$

*then $\alpha$ is irrational.*

There is a procedure that finds a good sequence to approximate any real number. Recall the greatest integer function $[x]$ = the largest integer $\leqslant x$.

Given a real number $\alpha$, consider the following iterative procedure:

$$\alpha = a_1 + \frac{1}{\alpha_2}, \quad \text{where } a_1 = [\alpha] \text{ and, if } a_1 \neq \alpha, \text{ then } 0 \leqslant \frac{1}{\alpha_2} \leqslant 1;$$

$$\alpha_2 = a_2 + \frac{1}{\alpha_3}, \quad \text{where } a_2 = [\alpha_2] \text{ and, if } a_2 \neq \alpha_2, \text{ then } 0 \leqslant \frac{1}{\alpha_3} \leqslant 1.$$

When $\alpha$ is rational, this procedure stops when $a_{n-1} = \alpha_{n-1}$. Otherwise, we continue, and we have

$$\alpha_n = a_n + \frac{1}{\alpha_{n+1}}, \quad \text{where } a_n = [\alpha_n] \text{ and so } 0 < \frac{1}{\alpha_{n+1}} \leqslant 1.$$

When the number $\alpha$ is irrational, this procedure never ends, and it produces a series of positive integers that can be turned into rational approximations to the number $\alpha$, through the continued fraction expansion of the number. When we substitute the $\alpha_i$ back into the procedure above we get the following expression, which defines the

compact notation for continued fractions:

$$\alpha = a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{a_4 + \cdots}}} = [a_1, a_2, a_3, a_4, \ldots].$$

We define the convergents of a continued fraction representation as the rational numbers

$$\frac{p_n}{q_n} = [a_1, a_2, \ldots, a_n].$$

For example, we calculate

$$\frac{15}{11} = 1 + \cfrac{1}{\cfrac{11}{4}} = 1 + \cfrac{1}{2 + \cfrac{1}{\cfrac{4}{3}}} = 1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{3}}} = [1, 2, 1, 3].$$

A more interesting example is given by $x = [1, 2, 2, 2, \ldots]$. Notice that

$$x = 1 + \frac{1}{[2, 2, 2, \ldots]} = 1 + \frac{1}{1 + [1, 2, 2, 2, \ldots]} = 1 + \frac{1}{1 + x}.$$

This leads to the equation $(x - 1)(x + 1) = 1$, or $x = \sqrt{2}$.

The convergents of a continued fraction enjoy some marvelous properties. We mention a few in the following proposition. For proofs, the reader can consult [**5**] or [**6**].

PROPOSITION 4. (1) *The convergents of a continued fraction satisfy the following recursion*:

$$p_1 = a_1, \qquad q_1 = 1,$$
$$p_2 = a_2 a_1 + 1, \qquad q_2 = a_2,$$

*and*

$$p_n = a_n p_{n-1} + p_{n-2}, \qquad q_n = a_n q_{n-1} + q_{n-2}, \qquad \textit{for } n \geqslant 3.$$

(2) *Furthermore*, $p_n q_{n-1} - p_{n-1} q_n = (-1)^n$ *for* $n > 1$, *from which it follows that*

$$\frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}} = \frac{(-1)^n}{q_{n-1} q_n}.$$

Notice that $p_{2n}/q_{2n} > p_{2n-1}/q_{2n-1}$ and $p_{2n+1}/q_{2n+1} < p_{2n}/q_{2n}$, since $q_m > 0$ for all $m$. It can be shown that the odd convergents form an increasing sequence, the even convergents a decreasing sequence, and for an infinite continued fraction, the values converge to an unique limit. The speed of the convergence is the most interesting feature of this construction.

THEOREM 5. *If* $\alpha = [a_1, a_2, a_3, \ldots]$, *and* $p_n/q_n = [a_1, a_2, \ldots, a_n]$, *then*

$$\frac{1}{2 q_n q_{n+1}} < \left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{q_n q_{n+1}}.$$

In the case that the sequence of $q_i$ is increasing, then it is the case that we have produced a collection of infinitely many distinct rationals, $p/q$, satisfying

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^2},$$

that is, any infinite continued fraction represents an irrational number.

**4. Newton's Method.** Having established some properties of real numbers that allow us to recognize irrational numbers, let's return to Fermat's Last Theorem. The next step is to determine the properties of $\sqrt[n]{1 - (p^n/q^n)}$. A natural way to obtain an approximation to $\sqrt[n]{1 - (p^n/q^n)}$ is via Newton's method. Define the function

$$f(x) = x^n - \left(1 - \frac{p^n}{q^n}\right).$$

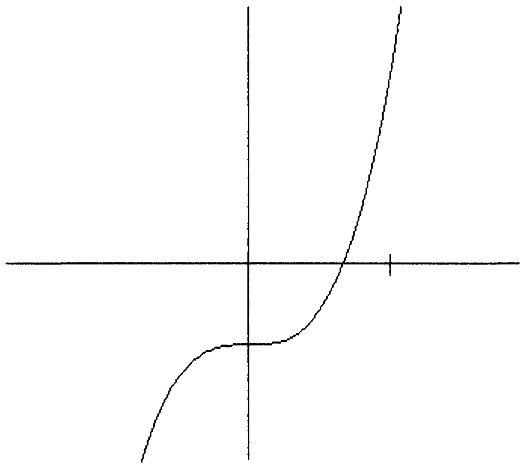


FIG. 2

Then we obtain a sequence converging to $\sqrt[n]{1 - (p^n/q^n)}$ by the recursion

$$x_0 = 1,$$

$$x_{i+1} = x_i - \frac{x_i^n - \left(1 - \dfrac{p^n}{q^n}\right)}{n x_i^{n-1}}.$$

It is clear that the real numbers $x_i$ are rational for all $i$, and as the graph in figure 2 shows, the function $f(x)$ is convex up over the region of interest and so

$$\lim_{i \to \infty} x_i = \sqrt[n]{1 - (p^n/q^n)}.$$

**Main Idea.** If, for $n \geqslant 3$ and for any $0 < p/q < 1$, the sequence of rationals produced by Newton's Method, $x_i = r_i/s_i$, satisfies

$$\left| \sqrt[n]{1 - \frac{p^n}{q^n}} - \frac{r_i}{s_i} \right| < \frac{1}{s_i^{1+\varepsilon}}$$

for some fixed $\varepsilon = \varepsilon(p/q) > 0$ and infinitely many values of $i$, then Fermat's Last Theorem holds for $n$.

Before developing the details of the Main Idea, let's consider an example that demonstrates how the Main Idea might work.

EXAMPLE. Let $\alpha = \sqrt{2}$. Newton's method considers the function $f(x) = x^2 - 2$ and so we get the recursion

$$x_0 = 1,$$
$$x_{i+1} = x_i - \frac{x_i^2 - 2}{2x_i} = \frac{x_i}{2} + \frac{1}{x_i}.$$

The sequence, $\{x_i\}$, consists entirely of rationals, and clearly converges to $\sqrt{2}$. There is another sequence of rationals converging to $\sqrt{2}$ given by the convergents of the continued fraction expansion of $\sqrt{2}$,

$$\frac{p_i}{q_i} = \Big[1, \underbrace{2, 2, \dots, 2}_{i}\Big].$$

The following observation gives an example of the Main Idea at work.

PROPOSITION 6. *For all $k \geqslant 0$, $x_k = p_{2^k-1}/q_{2^k-1}$.*

*Proof.* Introduce the function $F_k(x) = [1, \underbrace{2, 2, \dots, 2}_{k-1}, 1 + x]$. It follows immediately that $F_k(1) = p_k/q_k$, and furthermore, observe that

$$\frac{p_k}{q_k} = 1 + \cfrac{1}{1 + \cfrac{p_{k-1}}{q_{k-1}}} = \frac{p_{k-1} + 2q_{k-1}}{q_{k-1} + p_{k-1}}.$$

Now we can show that

$$F_k(x) = \frac{p_{k-1}x + 2q_{k-1}}{q_{k-1}x + p_{k-1}}.$$

For $k = 1$,

$$F_1(x) = 1 + \frac{1}{1 + x} = \frac{x + 2}{x + 1} = \frac{p_0 x + 2q_0}{q_0 x + p_0}.$$

By induction, we can consider

$$F_{k+1}(x) = 1 + \frac{1}{1 + F_k(x)} = 1 + \frac{1}{1 + \dfrac{p_{k-1}x + 2q_{k-1}}{q_{k-1}x + p_{k-1}}}$$

$$= \frac{(p_{k-1} + 2q_{k-1})x + 2(p_{k-1} + q_{k-1})}{(p_{k-1} + q_{k-1})x + (p_{k-1} + 2q_{k-1})} = \frac{p_k x + 2q_k}{q_k x + p_k}.$$

Finally, notice that $x_0 = 1 = p_0/q_0$. To prove the proposition, we proceed inductively; assume that $x_k = p_{2^k-1}/q_{2^k-1}$, then we have

$$\frac{p_{2^{k+1}-1}}{q_{2^{k+1}-1}} = \Big[1, \underbrace{2, 2, \ldots, 2}_{2^{k+1}-1}\Big]$$

$$= \Big[1, \underbrace{2, \ldots, 2}_{2^k - 1}, \underbrace{2, \ldots, 2}_{2^k}\Big] = \Big[1, \underbrace{2, \ldots, 2}_{2^k - 1}, 1 + \Big[1, \underbrace{2, \ldots, 2}_{2^k - 1}\Big]\Big]$$

$$= F_{2^k}\left(\frac{p_{2^k-1}}{q_{2^k-1}}\right) = \frac{p_{2^k-1}\dfrac{p_{2^k-1}}{q_{2^k-1}} + 2q_{2^k-1}}{q_{2^k-1}\dfrac{p_{2^k-1}}{q_{2^k-1}} + p_{2^k-1}}$$

$$= \frac{p_{2^k-1}^2 + 2q_{2^k-1}^2}{2p_{2^k-1}q_{2^k-1}} = \frac{p_{2^k-1}}{2q_{2^k-1}} + \frac{q_{2^k-1}}{p_{2^k-1}}$$

$$= \frac{x_k}{2} + \frac{1}{x_k} = x_{k+1}.$$

This proposition shows that Newton's method provides a subsequence of the continued fraction convergents of $\sqrt{2}$. Since $p_{n+1}/q_{n+1} = (p_n + 2q_n)/(q_n + p_n)$ it follows that $(p_n, q_n) = 1$ for all $n$ and that $q_{n+1} > q_n$. Thus we have proved the following.

COROLLARY 7. *The sequence of rational numbers $x_i = r_i/s_i$, obtained by Newton's method, converging to $\sqrt{2}$ from $x_0 = 1$, satisfies $|\sqrt{2} - r_n/s_n| < 1/s_n^2$, for all $n \geqslant 1$.*

The phenomenon explored in the example above is not particular to $\sqrt{2}$. The relationship between quadratic irrationals, Newton's Method and continued fraction convergents has been studied thoroughly (see, for example, [**4**], and the references there).

This example leads us on with the Main Idea, optimistically. For the general case, however, we need much closer scrutiny of the convergence of Newton's method.

**5. Convergence of Newton's method.** What follows is a review of the classical ideas that determine the speed of the convergence of Newton's Method, based on the discussion in [**10**]. Our focus is on the function $f(x) = x^n - (1 - p^n/q^n)$. Newton's Method is an iterative procedure and it is expressible in terms of a process in which a single function is iterated and converges to a fixed point. That function is

given by

$$G(x) = x - \frac{f(x)}{f'(x)},$$

away from the zeros of $f'(x)$, and the terms in the sequence produced by Newton's method are given by $x_{i+1} = G(x_i)$. In our case, the convergence of the sequence is given by

$$\lim_{i\to\infty} \underbrace{G \circ G \circ \cdots \circ G}_{i}(x) = \alpha = \sqrt[n]{1 - \frac{p^n}{q^n}}\,.$$

To study the convergence of Newton's Method we expand $G(x)$ into a Taylor polynomial around the value $\alpha$;

$$G(x) = G(\alpha) + (x - \alpha)G'(\alpha) + \frac{(x-\alpha)^2}{2}G''(\xi),$$

for some $\xi$, with $\alpha < \xi < x$. Now $G(\alpha) = \alpha$, and furthermore, we have

$$G'(\alpha) = 1 - \frac{(f'(\alpha))^2 - f(\alpha)f''(\alpha)}{(f'(\alpha))^2} = \frac{f(\alpha)f''(\alpha)}{(f'(\alpha))^2} = 0.$$

This simplifies the Taylor polynomial considerably. Likewise, when we compute the second derivative of $G$ at $\alpha$ we get

$$G''(\alpha) = \frac{(f'(\alpha))^2 f'(\alpha)f''(\alpha) + (f'(\alpha))^2 f(\alpha)f'''(\alpha) - 2f(\alpha)f'(\alpha)(f''(\alpha))^2}{(f'(\alpha))^4}$$

$$= \frac{f''(\alpha)}{f'(\alpha)}.$$

In our case, $f'(x) = nx^{n-1}$ and $f''(x) = n(n-1)x^{n-2}$, so we have that $G''(\alpha) = (n-1)/\alpha$. Since $n > 2$ in the case of interest for Fermat's Last Theorem and $0 < \alpha < 1$, we conclude that $G''(\alpha) > 2$.

Since $x_{i+1} = G(x_i)$, all of the above can be applied to the Taylor polynomial and so we get

$$x_{i+1} = \alpha + \frac{(x_i - \alpha)^2}{2}G''(\xi),$$

for some $\alpha < \xi < x_i$. Taking $i$ large enough, we have that $G''(\xi) > 2$ and so we conclude the following

PROPOSITION 8. *For $i$ large enough, the approximants produced by Newton's Method satisfy*

$$x_{i+1} - \alpha > (x_i - \alpha)^2.$$

The convergence of the sequence is seen here to be quadratic on the nose, however, the coefficient is not adding speed to the convergence but actually is stealing from it. The relation of this proposition to the actual terms in the sequence is exploited to prove the main theorem in the next section.

**6. Denominators and the Main Theorem.** Let us now consider the explicit formula for Newton's Method, in order to determine if we get the kind of approximation that is desired. The following proposition describes the behavior of the denominators produced in the process.

PROPOSITION 9. *If* $x_i = r_i/s_i$, $x_0 = 1$, *are the rationals converging to* $\sqrt[n]{1-(p/q)^n}$ *via Newton's Method, then, for i large,*

$$s_{i+1} > s_i^{n-3}.$$

*Proof.* We work a little more generally than we need to, in order to simplify the notation. Suppose $k/l$ is a rational that lies between 0 and 1, and is greater than $\sqrt[n]{1-(p/q)^n}$, like all of the approximants that Newton's Method will produce. We may suppose further that $l > q^n > n$, which is the case when $i$ is large for the Newton approximants $x_i$, since the function $f(x)$ is convex and nonlinear over the interval $(0,1)$ and so Newton's Method determines infinitely many distinct rationals enroute to the limit of the process; in particular, this condition holds for $i$ large if $k/l = x_i$.

Recall that $G(x) = x - f(x)/f'(x)$ with

$$f(x) = x^n - \left(1 - \left(\frac{p^n}{q^n}\right)\right) = \frac{q^n x^n - (q^n - p^n)}{q^n}.$$

Let $K/L = G(k/l)$ with $K/L$ in lowest terms. Explicitly we find that

$$G\left(\frac{k}{l}\right) = \frac{k}{l} - \frac{q^n\left(\frac{k}{l}\right)^n - (q^n - p^n)}{nq^n\left(\frac{k}{l}\right)^{n-1}} = \frac{k}{l} - \frac{q^n k^n - (q^n - p^n)l^n}{nq^n k^{n-1} l}$$

$$= \frac{nq^n k^n - q^n k^n + (q^n - p^n)l^n}{nq^n k^{n-1} l} = \frac{(n-1)q^n k^n + (q^n - p^n)l^n}{nq^n k^{n-1} l}.$$

Suppose the numerator and denominator have greatest common denominator $M$, then we have

$$K = \frac{(n-1)q^n k^n + (q^n - p^n)l^n}{M},$$

$$L = \frac{nq^n k^{n-1} l}{M}.$$

Suppose that $P$ is a prime that divides $M$. Then $P$ divides $nq^n k^{n-1} l$ and so either $P|n$, $P|q$, $P|k$, or $P|l$. These possibilities are not unrelated. Observe that

i) If $P|k$, then $P|q^n - p^n$, since $P$ does not divide $l$

ii) If $P|q$, then since $P$ does not divide $p$, $P$ does not divide $q^n - p^n$ and so $P$ divides $l$.

iii) If $P|l$, then either $P|q$, or $P|n-1$, since $P$ does not divide $k$.

From the observations given above, we may write $M = I_n I_k I_l$, where $I_n = (M, n)$, $I_k = (M, k^{n-1})$ and $I_l$ is $(M, q^n l)$. We conclude that

$$L > K = \frac{(n-1)q^n k^n + (q^n - p^n)l^n}{M}$$
$$= \frac{(n-1)q^n k^n}{I_n I_k I_l} + \frac{(q^n - p^n)}{I_n I_k I_l} l^n$$
$$= \frac{(n-1)q^n k^n}{I_n I_k I_l} + \frac{(q^n - p^n)}{I_k} \frac{l^n}{I_n I_l} > \frac{l^n}{I_n I_l}.$$

Since $I_n I_l$ divides $nq^n l$, we can conclude from $l > q^n > n$ that $l^n / I_n I_l > l^{n-3}$, and the proposition is proved.

We are now in a position to consider the Main Idea in some detail. Suppose that $\alpha = \sqrt[n]{(1 - p^n/q^n)}$, and that $x_i = r_i/s_i$ denotes the approximants due to Newton's Method. Suppose that for some $i$ large enough and for some $\varepsilon > 0$ we have the desired condition

$$\left| \alpha - \frac{r_i}{s_i} \right| < \frac{1}{s_i^{1+\varepsilon}}.$$

This can be written

$$\left| \alpha - \frac{r_i}{s_i} \right| = \frac{1}{s_i^{1+\eta}} < \frac{1}{s_i^{1+\varepsilon}},$$

for some $\eta > \varepsilon$.

Moving further on in the sequence, we may consider $|\alpha - r_{i+K}/s_{i+K}|$. From Proposition 8 we conclude the following string of inequalities:

$$\left| \alpha - \frac{r_{i+K}}{s_{i+K}} \right| > \left| \alpha - \frac{r_{i+K-1}}{s_{i+K-1}} \right|^2 > \left| \alpha - \frac{r_{i+K-2}}{s_{i+K-2}} \right|^4 > \cdots > \left| \alpha - \frac{r_{i+1}}{s_{i+1}} \right|^{2^{K-1}}$$
$$> \left| \alpha - \frac{r_i}{s_i} \right|^{2^K} = \frac{1}{s_i^{(1+\eta)2^K}}.$$

If we also have the desired condition,

$$\left| \alpha - \frac{r_{i+K}}{s_{i+K}} \right| < \frac{1}{s_{i+K}^{1+\varepsilon}},$$

then we can apply Proposition 9 to obtain another string of inequalities

$$\left| \alpha - \frac{r_{i+K}}{s_{i+K}} \right| < \frac{1}{s_{i+K}^{1+\varepsilon}} < \frac{1}{\left(s_{i+K-1}^{n-3}\right)^{1+\varepsilon}} < \cdots < \frac{1}{s_i^{(1+\varepsilon)(n-3)^K}}.$$

Putting the inequalities together we obtain

$$\frac{1}{s_i^{(1+\eta)2^K}} < \left| \alpha - \frac{r_{i+K}}{s_{i+K}} \right| < \frac{1}{s_i^{(1+\varepsilon)(n-3)^K}}.$$

Now, if $n > 5$, then there is some $K_0$ so that for all $K \geqslant K_0$ it follows that

$(1 + \varepsilon)(n - 3)^K > (1 + \eta)2^K$, and so our desired inequality leads to a contradiction. Thus we have proven

THEOREM 10. *If $n > 5$ and for some $i$ and for some $\varepsilon > 0$, we have that $|\alpha - r_i/s_i| < 1/s_i^{1+\varepsilon}$, then there is some $K_0$ so that, for all $K \geqslant K_0$,*

$$\left|\alpha - \frac{r_{i+K}}{s_{i+K}}\right| \geqslant \frac{1}{s_{i+K}^{1+\varepsilon}}.$$

From the theorem, we see that Newton's Method can produce only finitely many rationals that satisfy our desired condition for any given $\varepsilon > 0$. Thus the theorem condemns the Main Idea to the fate of most naive ideas—a place amid the failed attempts. It remains conceivable, however, that another numerical method could be applied in the same fashion and produce a proof of Fermat's Last Theorem following the outline given here. In fact, several results on the irrationality of classes of interesting numbers have been obtained in a similar fashion using Padé methods of approximation (see the work of Apéry, A. Baker, the Chudnovskys and Beukers, as surveyed in [**1**]). The work of Dyson [**2**] and Roth [**9**] reveals how subtle approximations of the kind sought here must be. For a review of recent progress on Fermat's Last Theorem, see [**8**].

**Acknowledgements.** This idea has been with me for a few years during which time I was encouraged by several people to see it through. In particular, John Ewing, S. J. Patterson, and John M. Mack offered helpful suggestions without which this article would have remained idle speculation. I owe them many thanks. I also thank Jeff Lagarias for an annotated bibliography of papers on the use of approximation methods in number theory, and the referee for ideas toward a better introduction.

REFERENCES

1. F. Beukers, Padé-approximations in number theory, in Padé-Approximation and Its Applications, Amsterdam 1980, Springer LNM 888 (1981), pp. 90–99.
2. F. J. Dyson, The approximation to algebraic numbers by rationals, *Acta Math.*, 79 (1947) 225–240.
3. H. M. Edwards, Fermat's Last Theorem; A Genetic Introduction to Algebraic Number Theory, Springer-Verlag, New York, 1977.
4. E. Frank and A. Sharma, Continued fraction expansions and iterations of Newton's method, *J. Reine Angew. Math.*, 219 (1965) 62–66.
5. S. Lang, Introduction to Diophantine Approximations, Addison-Wesley, Reading, MA, 1966.
6. C. D. Olds, Continued Fractions, MAA New Mathematical Library, Random House, 1963.
7. P. Ribenboim, 13 Lectures on Fermat's Last Theorem, Springer-Verlag, New York, 1979.
8. ______, Recent results about Fermat's last theorem, *Expositiones Math.*, 5 (1987) 75–90.
9. K. F. Roth, Rational approximations to algebraic numbers, *Mathematika*, 2 (1955) 1–20.
10. R. Wait, The Numerical Solutions of Algebraic Equations, John Wiley & Sons, New York, 1979.

# The Editor's Corner: Summability Theory: a Neglected Tool of Analysis

LEE A. RUBEL
*Department of Mathematics, University of Illinois, Urbana, IL 61801*

These days, the student of analysis usually learns at most three things about "summability theory" (="divergent series"). The first is that the Fourier series of an $L^1$ function is Cesàro summable to the function almost everywhere. The second (in a course in functional analysis) is the Silverman-Toeplitz criterion that characterizes "regular" methods of summability. The third, for the advanced student (in a course in Banach algebras), is Wiener's Tauberian theorem, but usually so dressed up that its applications to divergent series are lost in the shuffle. It is too bad that so little is taught about summability theory, since it remains a valuable tool in the armamentarium of the practicing analyst. In this note, I will point out three honest applications in the hope of rekindling some interest in the subject. Of course, the classic work in the field is Hardy's book [**3**]. As an afterthought, I will mention the paper [**2**] which applies summability theory to mathematical logic!

First of all, let us consider Borel summability. We say that $s_n \xrightarrow{B} s$ ($s_n$ converges to $s$ in the sense of Borel) if

$$\lim_{t \to +\infty} e^{-t} \sum_{n=0}^{\infty} \frac{s_n}{n!} t^n = s.$$

We stress that, throughout this note, we consider *complex-valued* sequences $(s_n)$. Two important facts are easy to prove, and are left to the reader as exercises.

(1) The Borel method is completely regular. That is, if $s_n \to s$ (including $s = +\infty$), then $s_n \xrightarrow{B} s$.

(2) If $s_n(z) = (z^{n+1} - 1)/(z - 1)$ is the *nth* partial sum of the geometric series $\sum_{n=0}^{\infty} z^n$, then $s_n(z) \xrightarrow{B} 1/(1 - z)$ in the half-plane Re $z < 1$.

It is not hard to show, using Cauchy's theorems, that if $f(z) = \sum_{n=0}^{\infty} a_n z^n$ is analytic in a neighborhood of the origin, and if $s_n(z)$ is the $n$th partial sum of the series then $s_n(z) \xrightarrow{B} f(z)$ in the Borel star of the region of analyticity of $f(z)$. This means, essentially, that through every singularity of $f(z)$ one draws the perpendicular to the ray joining that singularity to the origin, and considers the open half-plane, bounded by this perpendicular, that contains the origin. The Borel star is the intersection of all these half-planes, and is consequently a convex open set that contains the origin. In particular, the Taylor series of $f$ is Borel summable to $f$ in some neighborhood of every regular point of $f$ on the circle of convergence of the series.

We are now in a position to give a conceptual proof of Pringsheim's theorem (see [**4**, §7.21]), which says that if $f(z) = \sum_{n=0}^{\infty} a_n z^n$ has *positive* Taylor coefficients, then the real positive point on the circle of convergence must be a singularity of $f(z)$. For suppose it were not. Then by the above, the Taylor series would be Borel summable to $f(z)$ in a neighborhood of that point. But since the series diverges to $+\infty$ on the real axis beyond this point, and since Borel summability is completely regular, we have the impossible conclusion that $f$ is infinite beyond this point.

The next result we can use Borel summability for is the Hadamard gap theorem (see [**4**, §7.43]), which says that if $f(z) = \sum_{k=0}^{\infty} a_k z^k$, and $a_k = 0$ for $k \neq n_l$,

$l = 1, 2, \ldots,$ where $n_{l+1}/n_l \geqslant \theta$ for all $l$, for some number $\theta > 1$ (like $\Sigma z^{2^k}$), then every point of the circle of convergence of $f$ is a singular point of $f$—i.e., the circle of convergence is a "natural boundary."

More generally, we can prove Ostrowski's theorem on overconvergence (see [4, §7.41]), following a method of Zygmund (see [5]). We say that a power series $\Sigma a_n z^n$ has a gap $(n_k, n'_k)$ if $a_n = 0$ for $n_k < n < n'_k$. Ostrowski's theorem says that if $\lambda > 0$ and $\Sigma a_n z^n$ has an infinity of gaps satisfying $n'_k/n_k \geqslant 1 + \lambda$ and if its sum $f(z)$ has an analytic continuation to a neighborhood of a point $z_0$ on the circle of convergence, then the partial sums $s_{n_k}(z)$ converge uniformly to $f(z)$ in a neighborhood of $z_0$. For, under the hypotheses, $\Sigma a_k z^k$ is uniformly Borel summable to $f(z)$ in a neighborhood of $z_0$. We may suppose $z_0 = 1$, so that $a_0 + a_1 + a_2 + \cdots + a_n = O((1 + \delta)^n)$ for every positive $\delta$. Thus, we can apply the following "high-indices theorem" for Borel summability, whose proof is given in [5].

THEOREM. *Suppose that* $\lambda > 0$. *Then there is a number* $\delta = \delta(\lambda) > 0$ *such that, if*
(i) $\Sigma a_n$ *has an infinity of gaps* $(n_k, n'_k)$ *for which* $n'_k/n_k \geqslant 1 + \lambda$
(ii) $a_0 + a_1 + a_2 + \cdots + a_n = O((1 + \delta)^n)$
(iii) $\Sigma a_n$ *is Borel summable to* $s$,
*then*
(iv) $s_{n_k}$ *converges to* $s$ *where* $s_{n_k}$ *is the* $n_k$*th partial sum of* $\Sigma a_l$.

Notice that Hadamard's theorem follows from Ostrowski's, since in the Hadamard case, the sequence $s_{n_k}(z)$ is the *full* sequence of partial sums of $\Sigma a_n z^n$, so that if $f(z)$ has a point of regularity on the circle of convergence, then its power series would converge at a point *outside* the circle of convergence, which is manifestly impossible.

For the final result, we follow the approach of R. C. Buck in [1], and introduce the more powerful method of Mittag-Leffler summability. We say that $ML\ \Sigma_{n=0}^{\infty} a_n = s$ ("$\Sigma a_n$ is Mittag-Leffler summable to $s$") to mean that

$$\lim_{\alpha \to 0+} \sum_{n=0}^{\infty} \frac{a_n}{(\alpha n)!} = s,$$

where $(\alpha n)!$ means $\Gamma(\alpha n + 1)$, where $\Gamma$ is the Euler gamma function. We say that an entire function $f(z)$ is of exponential type if $|f(z)| \leqslant Ae^{B|z|}$ for some finite numbers $A$ and $B$ and all $z$. We say that $f(z) \in K(c)$ if $f(z)$ is of exponential type and if, further,

$$|f(iy)| \leqslant ae^{c|y|}$$

for some finite constant $a$ and all real $y$, which says that $f(z)$ grows no faster than $\sin cz$ on the imaginary axis. For *polynomials* $f(z)$, we all know the Newton expansion

$$f(z) = \sum_{n=0}^{\infty} \binom{z}{n} \Delta^n f(0),$$

where $\Delta^n$ is the $n$th difference with spacing 1. This formula recovers $f$ from its values at the nonnegative integers. Simple examples show, however, that the series needn't even converge for entire functions $f(z)$. But Buck proves (we do not give the details here) that if $c < \pi$ and if $f(z) \in K(c)$, then the Newton expansion for $f$ is Mittag-Leffler summable to $f(z)$ for all $z$.

As a corollary, Buck gets a new proof of the celebrated theorem of Carlson (known to information-theorists as "the sampling theorem," and rediscovered periodically) that if $f(z) \in K(c)$, $c < \pi$, and $f(n) = 0$ for $n = 0, 1, 2, \ldots$, then $f(z) \equiv 0$. Notice that $\pi$ is the right constant here, since $\sin \pi z$ vanishes at the integers without vanishing identically. Notice also that Carlson's theorem only says that it is possible to recover a suitable $f$ from its values only at the integers, but it does not say *how* to do the recovering. Buck's theorem does say *how*, and it involves Mittag-Leffler summability in a compelling way.

These methods enabled Buck to prove many new theorems about entire functions. There are hundreds of other applications of summability methods to classical and modern analysis, but I will have to stop here, and hope that I have made my point.

REFERENCES

1. R. C. Buck, Interpolation series, *Trans. Amer. Math. Soc.*, 64(1948) 283–298.
2. K. J. Compton, Application of a Tauberian theorem to finite model theory, *Arch. Math. Logik Grundlag.*, 25(1985) 91–98.
3. G. H. Hardy, Divergent Series, Oxford, 1949.
4. E. C. Titchmarsh, The Theory of Functions, second edition, Oxford, 1939.
5. A. Zygmund, On a theorem of Ostrowski, *J. London Math. Soc.*, 6(1931) 162–3.

# LETTERS TO THE EDITOR

Editor:

Most treatments of partial fractions, **[1]** being a recent example, assume the irreducibility of the factors of the denominator. This would require that, working over the real field, $x^2 + 2x - 1$ be replaced by $(x + 1 + \sqrt{2})(x + 1 - \sqrt{2})$, hardly something to be desired. That this condition is unnecessary follows from the Lemma (using the notation of **[1]**):

*If $p$, $w$ and $f$ are elements of $R[x]$ and $w$ and $f$ have no common factor then there exist unique elements $s$ and $t$ of $R[x]$, with either $s = 0$ or* degree($s$) < degree($f$), *such that*

$$p = sw + tf. \tag{1}$$

The basic step of the partial fraction decomposition follows on dividing (1) by $f^m w$.

*Proof of Lemma.* Since $w$ and $f$ have no common factor we can use the Euclidean Algorithm to construct polynomials $g$ and $h$ such that $gw + hf = 1$. Then $p = pgw + phf$. We can use the division transformation to construct $s$ such that $pg = kf + s$, with either $s = 0$ or degree($s$) < degree($f$). Hence $p = sw + (ph + kw)f$, and (1) follows on setting $t = ph + kw$. For uniqueness, if $p = sw + tf = s'w + t's$ then $(s - s')w = (t' - t)f$. Since $f$ and $w$ have no common factor $f$ must divide $s - s'$. But each of $s$ and $s'$ is either zero or of degree less than that of $f$, so $s - s' = 0$, and hence also $t' - t = 0$.

This is really the same proof as in **[1]** but with no use of quotient fields.

If $f$ is of degree 1 or 2 one can replace the Euclidean Algorithm with simple manipulation of the coefficients. This method was presented in **[2]** in the hope of inducing students to program it. I still have an old Fortran program that does this and also integrates the resulting partial fractions.


REFERENCES

1. Dan Scott and Donald R. Peeples, A constructive proof of the partial fraction decomposition, *Amer. Math. Monthly*, 95 (1988) 651–653.
2. Warren B. Stenberg and Robert J. Walker, Calculus, A Computer Oriented Presentation, vol. 2, The Center for Research in College Instruction of Science and Mathematics (CRICISAM), Florida State University, Tallahassee, FL, 1969.



Robert J. Walker
201 Adeline Avenue
Pittsburgh, PA 15228

# UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

*In this department the* MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

## Conway's RATS and Other Reversals

RICHARD K. GUY

Most of us met, at an early age, the "reverse subtract, reverse add" algorithm:

$$\begin{array}{r} 742 \\ \underline{247} \\ 495 \\ \underline{594} \\ 1089 \end{array} \qquad \begin{array}{r} 551 \\ \underline{155} \\ 396 \\ \underline{693} \\ 1089 \end{array} \qquad \begin{array}{r} 372 \\ \underline{273} \\ 099 \\ \underline{990} \\ 1089 \end{array} \qquad \begin{array}{r} 980 \\ \underline{089} \\ 891 \\ \underline{198} \\ 1089 \end{array}$$

and it still makes an attractive beginning algebra exercise. Many of us have also heard of the Kaprekar process [3], which iterates "sort, reverse, subtract," arriving at certain cycles or "Kaprekar constants":

$$\begin{array}{r} 321 \\ \underline{123} \\ 198 \end{array} \qquad \begin{array}{r} 981 \\ \underline{189} \\ 792 \end{array} \qquad \begin{array}{r} 972 \\ \underline{279} \\ 693 \end{array} \qquad \begin{array}{r} 963 \\ \underline{369} \\ 594 \end{array} \qquad \begin{array}{r} 954 \\ \underline{459} \\ 495 \end{array} \qquad \cdots$$

$$\begin{array}{r} 7731 \\ \underline{1377} \\ 6354 \end{array} \qquad \begin{array}{r} 6543 \\ \underline{3456} \\ 3087 \end{array} \qquad \begin{array}{r} 8730 \\ \underline{0378} \\ 8352 \end{array} \qquad \begin{array}{r} 8532 \\ \underline{2358} \\ 6174 \end{array} \qquad \begin{array}{r} 7641 \\ \underline{1467} \\ 6174 \end{array} \qquad \cdots$$

These are not the same as "Kaprekar numbers," whose squares yield the original number on adding the first and last "halves" of their digits: $142857^2 = 20408122449$ and $20408 + 122449 = 142857$.

Such digital oddities do not usually have much real mathematics in them, although it is sometimes of interest to ask what happens in bases other than 10. See [**1**], and the references there; also [**2, 4, 5, 6**].

Here are three problems that have come to light recently, each of which can consume unlimited amounts of computer time, perhaps without revealing anything significant.

*Shyam Sunder Gupta, P50, Badhwar Park, Colaba, Bombay-400005, India* notes that

$$65 - 56 = 3^2, \quad 65 + 56 = 11^2$$

$$621770 - 77126 = 738^2, \quad 621770 + 77126 = 836^2$$

and asks if there are other numbers, perhaps infinitely many, such that both the sum and the difference of the number and its reversal are perfect squares.

*Mario Borelli and Cecil B. Mast, Department of Mathematics, University of Notre Dame, Indiana 46556* say that an integer $n > 0$ is **palindromic** if there is a base $b$, $2 \leqslant b \leqslant n/2$, such that the representation of $n$ in base $b$, is a palindrome, i.e., reads the same left to right as right to left.

For example

$$11_{10} = 1011_2 = 102_3 = 23_4 = 21_5 \text{ is } \textit{not} \text{ palindromic, while}$$
$$13_{10} = 111_3 \textit{ is} \text{ palindromic.}$$

Of course, all numbers $b^k \pm 1$, $k \geqslant 2$, are palindromic: this includes the Mesenne and Fermat numbers.

It's not hard to prove that

*Every nonpalindromic integer* $\geqslant 7$ *is prime.*

But the converse is false, as the example $13_{10}$ shows.

The first few nonpalindromic integers are

$$(4,6), 11, 19, 53, 79, 103, 137, 139, 149, 163, 167,$$
$$179, 223, 263, 283, 293, 311, 317, 347$$

**Question 1:** Can palindromic or nonpalindromic primes be otherwise characterized?
**Question 2:** What is the cardinality, or the density, of the set of palindromic primes? Of the set of nonpalindromic primes?

*John Conway, Department of Mathematics, Princeton University, Princeton, NJ 08544* has recently offered money for a proof or disproof of his RATS conjecture, but I'll put nothing in print, to avoid a repetition of the somewhat embarrassing publicity he recently received in the presses on both sides of the Atlantic.

RATS is (the iteration of) the algorithm "reverse, add, then sort." In implementing this, zeros are suppressed. For example

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 4 | 8 | 16 | 77 | 145 | 668 | 1345 | 6677 | 13444 |
| 1 | 2 | 4 | 8 | 61 | 77 | 541 | 866 | 5431 | 7766 | 44431 |
| 2 | 4 | 8 | 16 | 77 | 154 | 686 | 1534 | 6776 | 14443 | 57875 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 55778 | 133345 | 666677 | 1333444 | 5567777 | 12333445 | 66666677 |
| 87755 | 543331 | 776666 | 4443331 | 7777655 | 54433321 | 77666666 |
| 143533 | 676676 | 1443343 | 5776775 | 13345432 | 66766766 | 144333343 |

| | | | | |
|---|---|---|---|---|
| 133333444 | 556667777 | 1233334444 | 5566667777 | 12333334444 |
| 444333331 | 777766655 | 4444333321 | 7777666655 | 44443333321 |
| 577666775 | 1334434432 | 5677667765 | 13344334432 | 56776667765 |

and a divergent pattern of period 2, $5^2 6^n 7^4$, $123^{n+1}4^4$ $(n = 3, 4, 5, \ldots)$ has emerged, where we've used superscripts to denote repetitions of digits.

Conway's RATS conjecture is that no matter what (base 10) number you start with, you either eventually join this divergent pattern, or enter a cycle.

Curt McMullen, also at Princeton, has made extensive RATS calculations. Conway's conjecture is true for all numbers less than a hundred million. These numbers either join the standard divergent sequence, or enter one of the first four members of the infinite family of 2-cycles, $(1^{2n}7^n, 2^n8^{2n})$, or the first member of the infinite family of 3-cycles, $(1^{6n}3^n, 2^{5n}4^{2n}, 4^{3n}6^{4n})$, or are tributary to one of the six following cycles:

| | | | | | | |
|---|---|---|---|---|---|---|
| Example | 3 | 29 | 69 | 3999 | 6999 | 27888 |
| Cycle length | 8 | 18 | 2 | 2 | 14 | 2 |
| Least member | 11 | 1223 | 78 | 11127 | 11144445 | 11667 |

McMullen has pursued many larger numbers, and discovered many other cycles, a few of which turn out to be members of more infinite families. Table 1 gives the cycle length and least member of nineteen other cycles and five other infinite families of cycles, in addition to the six and two already mentioned.

TABLE 1. Nineteen more cycles and five more infinite families.

| Cycle length | Least members | |
|---|---|---|
| 2 | $1^52^367$ | |
| 4 | $1^34^65^36, 1^92^77, 1^86^87^4, 1^{17}2^{15}67,$ | $1^{4n+6}7^{11n+15}$ |
| 5 | | $1^22^63^{58}4^{20n+200}5^{11n+265}$ |
| 6 | $1^42^65^2, 1^52^46^{14}7$ | |
| 7 | | $1^{47n+32}2^{80n+56}$ |
| 9 | $1^{23}2^2, 1^{27}$ | |
| 10 | $1^{21}2^{24}, 1^23^{88}5^26^2$ | $1^{325n+158}2^{16n+8}$ |
| 11 | $1^928^2$ | |
| 12 | $12^{21}3, 23^78^39^6, 1^{53}2^63^24,$ | $1^42^{1146n+32}3^{219n+12}$ |
| 14 | $1^{73}2^{38}3^24^4$ | |
| 18 | $1^{26}, 1^{14}2^{38}5^6$ | |
| 24 | $1^24^{48}5^{46}$ | |

What about other bases? It may be that base 10 is peculiar. McMullen found no examples of divergence in bases 2 to 9 or in bases 11 to 18. Of course, any number in base 2, after sorting, is $2^n - 1$ for some $n$, and adding the reversal gives $2(2^n - 1)$, which sorts back to $2^n - 1$: every number in base 2 is immediately tributary to a 1-cycle.

In base 3, a sorted number is of shape $1^a2^b$. If $a \geqslant b$, this goes to $1^{2b}$, then $2^{2b}$, $1^22^{2b-1}$, $1^42^{2b-3}$, $1^82^{2b-7}$ with the number of ones doubling until it exceeds the number of twos. If $a < b$, $1^a2^b$ goes to $1^{2a}2^{b-a}$ and again the number of ones doubles until it exceeds the number of twos. The only occasion on which a string can lengthen is when it consists entirely of twos. It is not hard to show that every number is tributary to a $k$-cycle:

$$1^{2^{k-1}-2}, 2^{2^{k-1}-2}, 1^22^{2^{k-1}-3}, 1^42^{2^{k-1}-5}, \ldots, 1^{2^{k-2}}2^{2^{k-2}-1}$$

for some $k \geqslant 3$.

It seems likely that a proof can be given for base 4 that all numbers eventually cycle. Can this be done for any other bases? In Base 19, McMullen found the

divergent sequence

$$123^3 4^4 5^{12} 6^n 7^{40},\ 8^2 9^2 A^6 B^8 C^{n+5} D^{38},\ 123^3 4^4 5^{12} 6^{n+1} 7^{40}, \ldots$$

where $A = 10, \ldots$ and $n \geqslant 19$. He also found that the sequence starting with 1 diverges in bases 37 and 50, but not in other bases, 20 to 49.

REFERENCES

1. Klaus E. Eldridge and Seok Sagong, The determination of Kaprekar and loop convergence of all three-digit numbers, this MONTHLY, 95 (1988) 105–112.
2. H. Hasse and G. D. Pritchett, The determination of all four-digit Kaprekar constants, *J. Reine Angew. Math.*, 299/300 (1978) 113–124.
3. D. R. Kaprekar, Another solitaire game, *Scripta Math.*, 15 (1949) 244–245.
4. D. R. Kaprekar, Problems involving reversal of digits, *Scripta Math.* 19 (1953) 81–82.
5. Andrzej Mąkowski, On Kaprekar's "junction numbers," *Math. Student*, 34 (1966) 77; MR 36 #6340.
6. W. Sierpiński, Sur les itérations de certaines fonctions numériques, *Rend. Circ. Mat. Palermo* (2), 13 (1964) 257–262; MR 32 #4077.

# NOTES

EDITED BY DAVID J. HALLENBECK, DENNIS DETURCK, AND ANITA E. SOLOW

## The Problem of the Calissons

GUY DAVID
*École Polytechnique, Centre de Mathématiques, 91128, Palaiseau Cedex, France*

CARLOS TOMEI
*Departamento de Matemática, PUC/RJ, R. Marquês de São Vicente, 225, Rio de Janeiro, Brasil*

A *calisson* is a French sweet that looks like two equilateral triangles meeting along an edge. Calissons could come in a box shaped like a regular hexagon, and their packing would suggest an interesting combinatorial problem. Suppose a box with side of length $n$ is filled with sweets of sides of length 1. The long diagonal of each calisson in the box is parallel to one of three different lines, as in the picture.

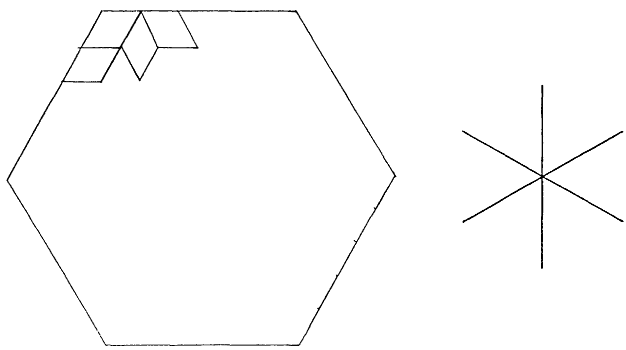


FIG. 1

We will refer to these three possibilities by saying that a calisson admits three distinct orientations. Our main result is the following: in any packing, the number of calissons with a given orientation is one third of the total number of calissons in the box.

C. Frederico Palmeira, of PUC-Rio, Brazil, heard of the problem from Jean Martinet (Université de Strasbourg), and showed it to one of the authors. We believe that the result is known and proven, but we were unable to find any references.

The idea of the proof is to reduce the problem to a very intuitive fact in three dimensions. We do not give complete details of a formal proof for two reasons. First, we do not want to spoil the simplicity of the basic intuitive idea. Furthermore, the result is an example of a proof by picture that does not translate immediately into precise mathematics. However, once some clues are given, we think it would be a nice exercise for an undergraduate class to fill in the details, by converting obvious geometric facts into precise mathematical statements.

We now sketch the proof. Draw an arbitrary filling of a box by calissons, such as the one below (in this case, $n$ is 5).

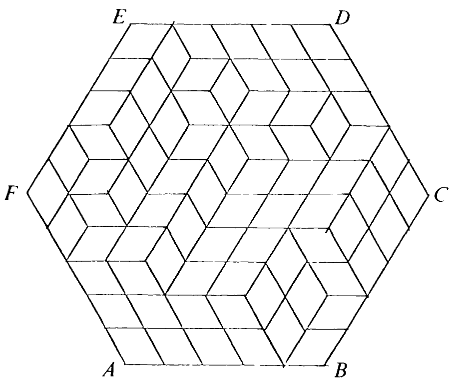



FIG. 2

If you rotate the picture counterclockwise a little bit, it is almost impossible not to 'see' a collection of cubes over a (square) floor, between two (square) walls.

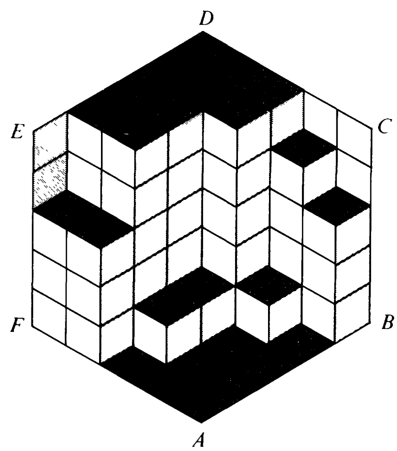



FIG. 3

If you were to look at this array of cubes from above, you would see exactly those faces with a certain orientation, and, of course, these faces will cover the floor. If you now look at a side wall, the same thing will happen—you will see all the faces with another fixed orientation covering the wall. So, the number of faces with a given orientation is the same for all three orientations, QED (informally)!

Let us give some idea of how a precise proof might go. First, it is not obvious that a filling of the hexagon corresponds to some arrangement of cubes within the walls and the floor. To prove this, we give a recipe for 'lifting' an arbitrary filling: we will assign coordinates to the vertices of the calissons as if they were points inside an $n \times n \times n$ cube in three-dimensional space. Give coordinates $(n, n, 0)$ to vertex $A$ in Figure 3. Now, join an arbitrary vertex $V$ to $A$ by a path along the sides of the calissons. Starting from $A$, coordinatize the vertices along the path according

to the following rules. Suppose you reach vertex $U$ by moving along a side of a calisson from vertex $W$ with coordinates $(a, b, c)$. Then:

a) if the side $WU$ is parallel to $AF$ in figure 3, give $U$ coordinates $(a, b - 1, c)$ if $U$ is higher than $W$ in figure 3 and coordinates $(a, b + 1, c)$ otherwise;

b) if $WU$ is parallel to $BC$, give $U$ coordinates $(a, b, c + 1)$ if $U$ is higher than $W$ and coordinates $(a, b, c - 1)$ otherwise;

c) if $WU$ is parallel to $AB$, give $U$ coordinates $(a - 1, b, c)$ if $U$ is higher than $W$ and coordinates $(a + 1, b, c)$ otherwise.

We leave to the reader the task of proving that the coordinatization above is well defined.

To prove the remaining statements in the informal proof, we give a convenient definition. Let $s$ be a side of a calisson along an edge of the box. An *s-chain* is a sequence of distinct calissons defined inductively as follows. The first calisson in the chain is the one having $s$ as a side. The $(k + 1)$st calisson meets the $k$th one at a side parallel to $s$. The edge $s$ can be parallel to three different directions, $d_1$, $d_2$ and $d_3$. We say that an s-chain is of *type* $d_i$ if $s$ is parallel to $d_i$. The following facts are rather easy to prove, using the coordinatization defined above.

1. An $s$-chain goes from one side of the box to the other side parallel to $s$.
2. The long diagonal of a calisson in an $s$-chain is never perpendicular to $s$.
3. Chains of different types have a calisson in common; chains of the same type never meet.
4. All calissons that lie in two chains of different type have the same orientation.

There are $n$ chains of each type. So, fixing two types, there are at least $n^2$ distinct calissons with a fixed orientation, from the facts above. As the total number of calissons is $3n^2$ (simply by dividing the area of the box by the area of one calisson), the number of calissons with a fixed orientation has to be $n^2$, and the main result is proved.

As a final exercise, we invite the reader to show that the configuration of calissons drawn below cannot be completed to fill a box (hint: notice that $n$ is 5, and one already has six cubes stacked one on top of the other).

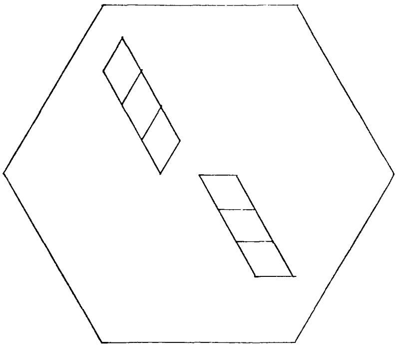


FIG. 4

The authors would like to thank the referees for suggesting many improvements in the presentation of the article.

# Arakelian's Approximation Theorem

JEAN-PIERRE ROSAY and WALTER RUDIN
*Department of Mathematics, University of Wisconsin, Madison, WI 53706*

Arakelian's theorem [**1**], [**3**] concerns uniform approximation by entire functions on possibly unbounded closed subsets $E$ of the complex plane $\mathbb{C}$. Our attention was drawn to this theorem because it has recently been used [**2**; p. 164], [**5**; p. 761] to construct interesting holomorphic maps from $\mathbb{C}^n$ to $\mathbb{C}^n$. The aim of our note is to show that Arakelian's theorem follows very easily from the much better known theorem of Mergelyan [6; Chap. 20] which deals with uniform approximation on *compact* sets.

Moreover—and this is perhaps our main point—in most applications the function that is to be approximated on $E$ is actually holomorphic in a *neighborhood* of $E$ (i.e., in an open set that contains $E$), and in that case the proof given below relies only on the classical approximation theorem of Runge [7], [**6**; p. 270]. (Incidentally, we recommend Runge's beautiful original paper; it is very readable.) For functions that are holomorphic in a neighborhood of $E$, Arakelian's theorem thus turns out to be really elementary.

If $E$ is a closed subset of $\mathbb{C}$, we shall use the phrase "hole of $E$" to denote any *bounded component of the complement of* $E$. Using this terminology, Runge's theorem states:

*If $K$ is a compact subset of $\mathbb{C}$, without holes, and $f$ is holomorphic in a neighborhood of $K$, then $f$ can be approximated, uniformly on $K$, by holomorphic polynomials.*

Mergelyan's theorem derives the same conclusion from a weaker assumption about $f$, namely: $f$ should be continuous on $K$ and holomorphic in the interior of $K$.

To motivate the definition that follows, note that if $E$ is a closed set without holes and $D$ is a closed disc in $\mathbb{C}$, then the intersection $E \cap D$ obviously has no holes either, but the union $E \cup D$ may very well have some, even infinitely many.

*Definition.* A closed set $E \subset \mathbb{C}$, without holes, is an *Arakelian set* if, for every closed disc $D \subset \mathbb{C}$, the *union* of all holes of $E \cup D$ is a bounded set.

*Note.* In [**1**] and [**3**], Arakelian's theorem is stated for closed sets without holes whose complement is "locally connected at infinity." The preceding definition describes the same class of sets. We chose it because we think that it is more easily understood, and because it explicitly states the property of $E$ that is crucial in our proof.

THEOREM. *If $E$ is an Arakelian set, $f$ is a complex-valued continuous function on $E$ that is holomorphic in the interior of $E$, and $\varepsilon > 0$, then there is an entire function $h$ that satisfies*

$$|h(z) - f(z)| < \varepsilon$$

*for every $z \in E$.*

The research for this paper was partially supported by NSF grant DMS-8400201 and by the William F. Vilas Trust Estate.

*Proof.* Since $E$ is an Arakelian set, there are closed discs $D_i$ $(i = 1, 2, 3, \ldots)$, centered at the origin, whose union is $\mathbb{C}$, so that the interior of $D_{i+1}$ contains the compact set $D_i \cup \overline{H}_i$, where $H_i$ is the union of the holes of $E \cup D_i$.

Put $E_0 = E$, and $E_i = E \cup D_i \cup \overline{H}_i$ for $i \geqslant 1$. Note that no $E_i$ has holes.

We first deal with the "Runge case" in which $f$ is holomorphic in a neighborhood of $E$. Put $h_0 = f$, fix $i \geqslant 1$, and assume (as induction hypothesis) that we have a function $h_{i-1}$ that is holomorphic in a neighborhood of $E_{i-1}$. There is an open disc $\Delta$ that contains $D_i \cup \overline{H}_i$ and whose closure $\overline{\Delta}$ lies in the interior of $D_{i+1}$. Choose a continuously differentiable function $\psi$ on $\mathbb{C}$ so that $0 \leqslant \psi \leqslant 1$, $\psi = 1$ in $\Delta$, $\psi = 0$ outside $D_{i+1}$.

Since $E_{i-1}$ has no holes, the same is true of $E_{i-1} \cap D_{i+1}$. Runge's theorem therefore furnishes a polynomial $P$ so that

$$|h_{i-1} - P| < 2^{-i-1}\varepsilon \quad \text{on } E_{i-1} \cap D_{i+1} \tag{1}$$

and

$$\frac{1}{\pi}\int_{E_{i-1}} |(h_{i-1} - P)(w)(\bar{\partial}\psi)(w)| \frac{dm(w)}{|z - w|} < 2^{-i-1}\varepsilon \tag{2}$$

for all $z \in \mathbb{C}$. Here $\bar{\partial}\psi = \partial\psi/\partial\bar{z}$, and $m$ denotes two-dimensional Lebesgue measure. Note that (2) can be achieved because the integrand vanishes outside $D_{i+1}$.

Now let $V$ be a neighborhood of $E_{i-1}$ in which $h_{i-1}$ is holomorphic, and which is so close to $E_{i-1}$ that (2) holds with $V$ in place of $E_{i-1}$. Define

$$r(z) = \frac{1}{\pi}\int_V (h_{i-1} - P)(w)(\bar{\partial}\psi)(w)\frac{dm(w)}{z - w} \qquad (z \in \mathbb{C}) \tag{3}$$

and

$$h_i = \psi P + (1 - \psi)h_{i-1} + r \text{ in } \Delta \cup V. \tag{4}$$

(This is well defined because $1 - \psi = 0$ in $\Delta$.)

The fact that $(h_{i-1} - P)\bar{\partial}\psi$ is continuously differentiable in $V$ shows that

$$\bar{\partial}r = (h_{i-1} - P)\bar{\partial}\psi$$

in $V$; see, for example [**4**; p. 104]. Since $\bar{\partial}h_{i-1} = 0$ in $V$, we see that

$$\bar{\partial}h_i = (P - h_{i-1})\bar{\partial}\psi + \bar{\partial}r = 0 \text{ in } V. \tag{5}$$

In $\Delta$, $\bar{\partial}\psi = 0$. The integral (3) extends therefore only over $V \setminus \Delta$, so that $r$ is holomorphic in $\Delta$. The same is true of $h_i$, because $h_i = P + r$ in $\Delta$.

We conclude: $h_i$ is holomorphic in the neighborhood $\Delta \cup V$ of $E_i$ (which is our induction hypothesis, with $i$ in place of $i - 1$) and

$$|h_i - h_{i-1}| = |(P - h_i)\psi + r| < 2^{-i}\varepsilon \text{ on } E_{i-1} \tag{6}$$

by (1); note that $\psi = 0$ outside $D_{i+1}$ and that $|r| < 2^{-i-1}\varepsilon$.

The sets $E_{i-1}$ contain the discs $D_{i-1}$, and these expand to cover $\mathbb{C}$. Hence (6) shows that $h = \lim h_i$ satisfies the conclusion of the theorem.

In the "Mergelyan case" we get $P$ from Mergelyan's Theorem, define $r(z)$ as above, but with $E_{i-1}$ in place of $V$ in (3), and we conclude that $h_i$ (defined by (4) on $\Delta \cup E_{i-1}$) is continuous on $E_i$, holomorphic in the interior of $E_i$, and satisfies (6) on $E_{i-1}$.

This concludes the proof.

*Remark.* On sets without interior, a considerably stronger version of the theorem can be derived from it without any extra effort:

*If $E$ is an Arakelian set with empty interior, $f$ and $\omega$ are continuous functions on $E$, $f$ is complex-valued, $\omega$ is positive (and $\omega(z) \to 0$ as $z \to \infty$ along $E$, to make things interesting), then there is an entire function $h$ that satisfies*

$$|h(z) - f(z)| < \omega(z)$$

*for every $z \in E$.*

To prove this, apply the theorem twice: There are entire functions $g$ and $h_0$ so that

$$\operatorname{Re} g < \log \omega \text{ and } |h_0 - f \cdot \exp(-g)| < 1$$

on $E$. Put $h = h_0 \cdot \exp(g)$.

REFERENCES

1. N. U. Arakelian, Uniform approximation on closed sets by entire functions, *Izv. Akad. Nauk SSSR*, 28 (1964) 1187–1206 (Russian).
2. P. G. Dixon and J. Esterle, Michael's problem and the Poincaré-Fatou-Bieberbach phenomenon, *Bull. Amer. Math. Soc.* (New Series), 15 (1986) 127–187.
3. W. H. Fuchs, Théorie de l'approximation des fonctions d'une variable complexe, Presses de l'Université de Montréal, 1968.
4. R. Narasimhan, Complex Analysis in One Variable, Birkhäuser, 1985.
5. Y. Nishimura, Applications holomorphes injectives de $\mathbb{C}^2$ dans lui-même qui exceptent une droite complexe, *J. Math. Kyoto Univ.* 4 (1984) 755–761.
6. W. Rudin, Real and Complex Analysis, 3rd ed., McGraw-Hill, 1987.
7. C. Runge, Zur Theorie der eindeutigen analytischen Funktionen, *Acta Math.* 6 (1885) 229–244.

Added in proof: Another reference to Arakelian's theorem is Approximation Uniforme Qualitative sur les Ensembles non Bornés, by P. M. Gauthier and W. Hengartner, Presses de l'Université de Montréal, 1962, p. 37.

## The Norm of a Linear Functional

I. J. MADDOX
*Department of Pure Mathematics, Queen's University, Belfast BT7 1NN, Northern Ireland*

Bounded linear functionals of the type

$$f(x) = \int_a^b g(t)x(t)\,dt \tag{1}$$

frequently occur in elementary functional analysis and its applications, and one needs to have an expression for $\|f\|$, the norm of $f$. For example, if $x = x(t)$ is a continuous function of period $2\pi$ and $X$ is the Banach space of all such functions, with $\|x\| = \max\{|x(t)|: -\pi \leqslant t \leqslant \pi\}$, then the Fourier coefficients of $x$ are, by definition,

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} x(t)\cos kt\,dt, \qquad b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} x(t)\sin kt\,dt, \tag{2}$$

for $k = 0, 1, 2, \dots$ . Now if $s_n(x)$ denotes the $n$th partial sum of the Fourier series

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kw + b_k \sin kw)$$

in the case when $w = 0$, we see from (2) that

$$s_n(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(1 + 2 \sum_{k=1}^{n} \cos kt\right) x(t)\, dt, \tag{3}$$

which is of the form (1), with $x$ a continuous function on $[-\pi, \pi]$ and with

$$g(t) = \frac{1}{2\pi}\left(1 + 2 \sum_{k=1}^{n} \cos kt\right) = \frac{1}{2\pi} \cdot \frac{\sin(n + 1/2)t}{\sin(t/2)}. \tag{4}$$

It follows from (3) and (4) that

$$|s_n(x)| \leqslant \frac{1}{2\pi}\left(1 + \sum_{k=1}^{n} 2\right) \int_{-\pi}^{\pi} |x(t)|\, dt \leqslant (1 + 2n)\|x\|,$$

which implies that $s_n$ is a bounded (and obviously linear) functional on $X$. Since, by definition, $\|s_n\| = \sup\{|s_n(x)|\colon \|x\| \leqslant 1\}$, it follows that $\|s_n\| \leqslant 1 + 2n$.

In order to deduce the very interesting result that there is a continuous function of period $2\pi$ whose Fourier series diverges at 0 it is enough to show that $\|s_n\| \to \infty$ as $n \to \infty$. Details may be found in Gál **[1]**, where it is shown that

$$\|s_n\| = \int_{-\pi}^{\pi} |g(t)|\, dt, \tag{5}$$

with $g$ given by (4).

Because of the especially simple form of $g$ in (5) the determination of $\|s_n\|$ is not difficult, but in the more general case (1), where $x$ may be any continuous function on $[a, b]$ and $g$ any given continuous function the usual proofs which determine $\|f\|$ are quite technical and reasonably lengthy. See, for example, Kantorovich and Akilov **[2]**, p. 104.

In this note we show how to determine $\|f\|$ in (1) by a short elementary method which holds also for any Riemann integrable functions.

Now let us denote by $R[a, b]$ the real linear space of Riemann integrable functions $x = x(t)$, $a \leqslant t \leqslant b$, with $a$ and $b$ finite real numbers. Define the norm of $x$ to be $\|x\| = \sup\{|x(t)|\colon a \leqslant t \leqslant b\}$. Since every continuous $x$ is in $R[a, b]$ we have that $C[a, b]$ is a linear subspace of $R[a, b]$ and so the theorem below also applies to any continuous functions.

THEOREM 1. *Let $g$ be a fixed element of $R[a, b]$. Then $f$ given by* (1) *above defines a bounded linear functional on $R[a, b]$ and*

$$\|f\| = \sup\{|f(x)|\colon \|x\| \leqslant 1\} = \int_a^b |g(t)|\, dt. \tag{6}$$

*Proof.* First, for any $x \in R[a, b]$ with $\|x\| \leqslant 1$ we have

$$|f(x)| \leqslant \int_a^b |g(t)| \cdot |x(t)|\, dt \leqslant \|x\| \int_a^b |g(t)|\, dt,$$

which implies $\|f\| \leqslant \int_a^b |g(t)|\,dt$. The problem is to prove the reverse inequality. To do this, take any natural number $n$. Then, writing for simplicity, $\int |g| = \int_a^b |g(t)|\,dt$, etc., we have

$$\int |g| = \int |g| \cdot \frac{1 + n|g|}{1 + n|g|}$$
$$= \int \frac{|g|}{1 + n|g|} + \int g \cdot \frac{ng}{1 + n|g|} = \int \frac{|g|}{1 + n|g|} + f\left(\frac{ng}{1 + n|g|}\right)$$
$$\leqslant \int \frac{dt}{n} + \|f\| \left\| \frac{ng}{1 + n|g|} \right\|,$$

using the facts that $|g(t)|/(1 + n|g(t)|) < 1/n$ and that $|f(x)| \leqslant \|f\|\,\|x\|$ for all $x \in R[a, b]$. Also, since $\|ng/(1 + n|g|)\| \leqslant 1$, we see that

$$\int_a^b |g(t)|\,dt \leqslant (b - a)/n + \|f\|, \tag{7}$$

and our result follows by letting $n \to \infty$ in (7).

An extension of Theorem 1 is obtained by considering the space $R_B[a, \infty)$ of Riemann integrable bounded functions $x$ on $[a, \infty)$ with $\|x\| = \sup\{|x(t)|\colon a \leqslant t < \infty\}$, and a fixed $g \in R[a, y]$ for each $y > a$ which is such that

$$\lim_{y \to \infty} \int_a^y |g(t)|\,dt = \int_a^\infty |g(t)|\,dt < \infty. \tag{8}$$

In this case,

$$f(x) = \int_a^\infty g(t)x(t)\,dt$$

defines a bounded linear functional $f$ on $R_B[a, \infty)$ such that

$$\|f\| = \int_a^\infty |g(t)|\,dt. \tag{9}$$

That $\|f\| \leqslant \int_a^\infty |g(t)|\,dt$ is trivial. To prove the reverse inequality we take any $\varepsilon > 0$. Then by (8) there exists $b = b(\varepsilon) > a$ such that

$$\int_a^\infty |g(t)|\,dt < \int_a^b |g(t)|\,dt + \varepsilon.$$

Hence (7) above implies

$$\int_a^\infty |g(t)|\,dt < (b - a)/n + \|f\| + \varepsilon,$$

and (9) follows on letting $n \to \infty$, and then letting $\varepsilon \to 0$.

My thanks are due to my colleague, Dr. David Armitage, for his response to a question of mine which led me to obtain the simple proof given above.

## REFERENCES

1. I. S. Gál, On sequences of operations in complete vector spaces, this MONTHLY, 60 (1953) 527–538.
2. L. V. Kantorovich and G. P. Akilov, Functional Analysis in Normed Spaces, Pergamon Press, New York, 1964.

# An Elementary Treatment of the Radon-Nikodym Derivative

Richard C. Bradley
*Department of Mathematics, Indiana University, Bloomington, Indiana 47405*

The purpose of this note is to review an elementary proof of the Radon-Nikodym theorem. The proof given here is similar to other well known ones, but has one extra little twist which seems to simplify the arithmetic and which apparently is not well known.

Our main tools will be the following:

(i) Elementary properties of abstract integration (Chapter 1 of Rudin **[2]**);
(ii) Elementary properties of partitions—for example, the "smallest" (i.e. least refined) common refinement of partitions $\{A_1, \ldots, A_I\}$, $\{B_1, \ldots, B_J\}$, and $\{C_1, \ldots, C_K\}$ of the same set $X$ is the partition $\{A_i \cap B_j \cap C_k\colon 1 \leqslant i \leqslant I, 1 \leqslant j \leqslant J, 1 \leqslant k \leqslant K\}$;
(iii) The Cauchy-Schwarz inequality

$$\int_X |fg|\, d\mu \leqslant \left[\int_X |f|^2\, d\mu\right]^{1/2} \left[\int_X |g|^2\, d\mu\right]^{1/2} \tag{1}$$

for measurable functions $f$ and $g$ on a measure space $(X, \mathscr{A}, \mu)$.

The Cauchy-Schwarz inequality has a well known simple proof (see e.g. Apostol **[1**, vol. 2, p. 394, exercise 4] for the case of Lebesgue measure).

The key to our development of the Radon-Nikodym theorem is the following special case:

Theorem 1 (Radon-Nikodym, simple case). *Suppose $\lambda$ and $\mu$ are positive measures on a measurable space $(X, \mathscr{A})$ such that*

(i) $\mu(X) = 1$ *and*
(ii) $\forall A \in \mathscr{A}$, $\lambda(A) \leqslant \mu(A)$.

*Then there exists an $\mathscr{A}$-measurable function $h$ on $X$, with $0 \leqslant h(x) \leqslant 1 \forall x \in X$, such that $\forall A \in \mathscr{A}$, $\lambda(A) = \int_A h\, d\mu$.*

Hypothesis (ii) is of course much stronger than the standard hypothesis that $\lambda$ is absolutely continuous with respect to $\mu$. Later on, we shall derive the standard Radon-Nikodym theorem as an easy consequence of the special case treated in Theorem 1.

*Proof of Theorem 1.*

(A) *Preliminaries*

Definition 1. *For each finite $\mathscr{A}$-measurable partition $\mathscr{P} = \{A_1, \ldots, A_n\}$ of $X$, define the function $h_{\mathscr{P}}$ on $X$ as follows*:

$$\forall k = 1, 2, \ldots, n, \qquad \forall x \in A_k,$$

$$h_{\mathscr{P}}(x) = \begin{cases} \dfrac{\lambda(A_k)}{\mu(A_k)} & \text{if } \mu(A_k) > 0 \\ 0 & \text{if } \mu(A_k) = 0. \end{cases} \tag{2}$$

(The phrase "$\mathscr{A}$-measurable partition" simply means that each $A_k$ in $\not{p}$ is a member of the $\sigma$-field $\mathscr{A}$.)

In the context of Definition 1, one has that

$$\forall x \in X, \qquad 0 \leqslant h_{\not{p}}(x) \leqslant 1. \tag{3}$$

LEMMA 1. *In the context of Definition* 1, *if* $A = \bigcup_{k \in I} A_k$ *for some index set* $I \subset \{1, 2, \ldots, n\}$, *then* $\lambda(A) = \int_A h_{\not{p}}\, d\mu$. *In particular*, $\lambda(X) = \int_X h_{\not{p}}\, d\mu$.

LEMMA 2. *Suppose* $\not{p}$ *and* $\mathscr{Q}$ *are finite* $\mathscr{A}$*-measurable partitions of* $X$, *with* $\mathscr{Q}$ *being a refinement of* $\not{p}$. *Then*

$$\int_X h_{\mathscr{Q}}^2\, d\mu = \int_X h_{\not{p}}^2\, d\mu + \int_X \left(h_{\mathscr{Q}} - h_{\not{p}}\right)^2 d\mu \geqslant \int_X h_{\not{p}}^2\, d\mu.$$

Lemma 1 is trivial. To prove Lemma 2 we use Lemma 1 and the trivial fact that for each partition $\not{p} = \{A_1, \ldots, A_n\}$ the function $h_{\not{p}}$ is constant on each $A \in \not{p}$ (i.e., on each of the sets $A_1, \ldots, A_n$). Here are the steps:

$$\forall A \in \not{p}, \qquad \int_A h_{\mathscr{Q}}\, d\mu = \lambda(A) = \int_A h_{\not{p}}\, d\mu.$$

$$\forall A \in \not{p}, \qquad \int_A h_{\not{p}} h_{\mathscr{Q}}\, d\mu = \int_A h_{\not{p}}^2\, d\mu \text{ and hence } \int_A h_{\not{p}}\left(h_{\mathscr{Q}} - h_{\not{p}}\right) d\mu = 0.$$

$\int_X h_{\not{p}}\left(h_{\mathscr{Q}} - h_{\not{p}}\right) d\mu = 0$, and Lemma 2 follows.

(B) *Construction of the function* $h$.

Define

$$c = \sup \int_X h_{\not{p}}^2\, d\mu, \tag{4}$$

where this sup is taken over all finite $\mathscr{A}$-measurable partitions $\not{p}$ of $X$. By eq. (3) and our hypothesis that $\mu(X) = 1$, we have $0 \leqslant c \leqslant 1$.

For each $n = 1, 2, 3, \ldots$ let $\not{p}_n$ be a finite $\mathscr{A}$-measurable partition of $X$ such that

$$\int_X h_{\not{p}_n}^2\, d\mu \geqslant c - (1/4)^n.$$

For each $n = 1, 2, 3, \ldots$ let $\mathscr{Q}_n$ be the "smallest" common refinement of the partitions $\not{p}_1, \not{p}_2, \ldots, \not{p}_n$. By Lemma 2,

$$\begin{aligned} \forall n \geqslant 1, \qquad c - (1/4)^n \leqslant \int_X h_{\not{p}_n}^2\, d\mu &\leqslant \int_X h_{\mathscr{Q}_n}^2\, d\mu \\ &\leqslant \int_X h_{\mathscr{Q}_{n+1}}^2\, d\mu \leqslant c. \end{aligned} \tag{5}$$

Hence, for each $n \geqslant 1$,

$$\int_X \left(h_{\mathscr{Q}_{n+1}} - h_{\mathscr{Q}_n}\right)^2 d\mu = \int_X h_{\mathscr{Q}_{n+1}}^2\, d\mu - \int_X h_{\mathscr{Q}_n}^2\, d\mu \leqslant (1/4)^n$$

by Lemma 2, and hence

$$\int_X |h_{\mathscr{Q}_{n+1}} - h_{\mathscr{Q}_n}|\, d\mu \leqslant (1/2)^n$$

by the Cauchy-Schwarz inequality (eq. (1) with

$$f = h_{\mathscr{Q}_{n+1}} - h_{\mathscr{Q}_n}$$

and $g = 1$) and our assumption that $\mu(X) = 1$. Hence

$$\int_X \left( \sum_{n=1}^{\infty} |h_{\mathscr{Q}_{n+1}} - h_{\mathscr{Q}_n}| \right) d\mu < \infty .$$

Hence

$$\sum_{n=1}^{\infty} |h_{\mathscr{Q}_{n+1}} - h_{\mathscr{Q}_n}| < \infty \quad \text{a.e. } -\mu$$

by a trivial property of integrals.

Hence

$$\lim_{n\to\infty} h_{\mathscr{Q}_n} \text{ exists a.e. } -\mu .$$

DEFINITION 2. *Define the function h on X by*

$$h(x) = \begin{cases} \lim_{n\to\infty} h_{\mathscr{Q}_n}(x) & \textit{if this limit exists} \\ 0 & \textit{otherwise}. \end{cases}$$

(C) *Verification that h has the right properties.*

The function $h$ is $\mathscr{A}$-measurable, and by eq. (3) one has that $0 \leqslant h(x) \leqslant 1$ for all $x \in X$. All that remains is to prove that $\forall A \in \mathscr{A}$, $\lambda(A) = \int_A h\, d\mu$.

Let $A \in \mathscr{A}$ be arbitrary but fixed. For each $n = 1, 2, 3, \ldots$ let $\mathscr{R}_n$ be the "smallest" common refinement of the partitions $\mathscr{Q}_n$ and $\{A, X - A\}$. For each $n \geqslant 1$,

$$c - (1/4)^n \leqslant \int_X h_{\mathscr{Q}_n}^2\, d\mu \leqslant \int_X h_{\mathscr{R}_n}^2\, d\mu \leqslant c$$

by Lemma 2 and equations (4) and (5), and hence

$$\int_X (h_{\mathscr{R}_n} - h_{\mathscr{Q}_n})^2\, d\mu \leqslant (1/4)^n$$

by Lemma 2. Hence by another application of the Cauchy–Schwarz inequality,

$$\forall n \geqslant 1, \qquad \left| \int_A (h_{\mathscr{R}_n} - h_{\mathscr{Q}_n})\, d\mu \right| \leqslant \int_A |h_{\mathscr{R}_n} - h_{\mathscr{Q}_n}|\, d\mu \leqslant (1/2)^n . \tag{6}$$

Now for each $n = 1, 2, 3, \ldots$ by Lemma 1,

$$\lambda(A) = \int_A h_{\mathscr{R}_n}\, d\mu = \int_A (h_{\mathscr{R}_n} - h_{\mathscr{Q}_n})\, d\mu + \int_A h_{\mathscr{Q}_n}\, d\mu . \tag{7}$$

By (6),

$$\lim_{n\to\infty} \int_A (h_{\mathscr{R}_n} - h_{\mathscr{Q}_n})\, d\mu = 0.$$

Also,

$$\lim_{n\to\infty}\int_A h_{\mathscr{Q}_n}\,d\mu = \int_A h\,d\mu$$

by Definition 2 and the dominated convergence theorem.

Hence by equation (7), $\lambda(A) = \int_A h\,d\mu$.

This completes the proof of Theorem 1.

We will need two elementary corollaries of Theorem 1.

COROLLARY 1. *The statement of Theorem* 1 *still holds if the hypothesis* $\mu(X) = 1$ *there is replaced by the hypothesis* $0 < \mu(X) < \infty$.

COROLLARY 2. *If* $\lambda$ *and* $\mu$ *and* $h$ *are as in Theorem* 1 *or Corollary* 1, *then the equation*

$$\int_X g\,d\lambda = \int_X gh\,d\mu$$

*holds for every real nonnegative* $\mathscr{A}$*-measurable function* $g$ *on* $X$, *and also for every complex-valued function* $g \in \mathscr{L}_1(\lambda)$.

Now we come to the standard form of the Radon-Nikodym theorem.

THEOREM 2 (Radon-Nikodym). *Suppose* $\lambda$ *and* $\mu$ *are positive measures on a measurable space* $(X, \mathscr{A})$ *such that* $0 < \lambda(X) < \infty$, $0 < \mu(X) < \infty$, *and* $\lambda$ *is absolutely continuous with respect to* $\mu$. *Then there exists a real nonnegative* $\mathscr{A}$*-measurable function* $h$ *on* $X$ *such that* $\forall A \in \mathscr{A}$, $\lambda(A) = \int_A h\,d\mu$.

*Proof.* By Corollary 1 there exist Radon-Nikodym derivatives

$$h_\lambda = \frac{d\lambda}{d(\lambda+\mu)} \quad\text{and}\quad h_\mu = \frac{d\mu}{d(\lambda+\mu)}.$$

Define $B = \{x \in X\colon h_\mu(x) > 0\}$ and $C = \{x \in X\colon h_\mu(x) = 0\}$. Now $\mu(C) = \int_C h_\mu\,d(\lambda+\mu) = 0$, and hence $\lambda(C) = 0$ (since $\lambda \ll \mu$). Define the function $h$ on $X$ by

$$h(x) = \begin{cases} \dfrac{h_\lambda(x)}{h_\mu(x)} & \text{if } x \in B \\ 0 & \text{if } x \in C. \end{cases}$$

If $A \in \mathscr{A}$ and $A \subset B$, then

$$\lambda(A) = \int_A h_\lambda\,d(\lambda+\mu) = \int_A h\cdot h_\mu\,d(\lambda+\mu) = \int_A h\,d\mu$$

by Corollary 2. Since $\lambda$ and $\mu$ are both 0 on $C$, one has that $\lambda(A) = \int_A h\,d\mu$ for all $A \in \mathscr{A}$. This completes the proof.

*Acknowledgments*: The author is indebted to Alberto Torchinsky, Bill Ziemer, and Boh Hrees for their encouragement; to Max Zorn for calling attention to an error of terminology in an earlier draft of this paper; and to a referee for his helpful suggestions.

REFERENCES

1. T. M. Apostol, Calculus, vol. 2, Blaisdell, Waltham, 1962.
2. W. Rudin, Real and Complex Analysis, 2nd ed., McGraw-Hill, New York, 1974.

EDITED BY MELVIN HENRIKSEN AND STAN WAGON

# On the Proof of the Radon-Nikodym Theorem

ALBERT WILANSKY
*Department of Mathematics, Lehigh University, Bethlehem, PA 18015*

The elegant proof (call it Pl) of the Radon-Nikodym Theorem, due to von Neumann and given, for example, in [**1**, Chap. 11, Ex. 37] does not give the result in as much generality as the much less pleasing proof of [**1**, Chap. 11, Sec. 6]. The purpose of this note is to derive the general result from von Neumann's, thus enabling the elegant proof Pl to be given without loss.

Proof Pl yields that if $\mu$, $\nu$ are $\sigma$-finite measures with $\nu \ll \mu$, then there exists $f$ such that $\nu(E) = \int_E f\,d\mu$ for all measurable $E$. This will now be extended so as to drop the assumption that $\nu$ is $\sigma$-finite.

LEMMA. *Let $(X, B, \mu)$ be a finite measure space and $\nu$ a measure with $\nu \ll \mu$. Then $X$ is a disjoint union $Y \cup Z$ such that $\nu$ is $\sigma$-finite on $Y$ and $\nu(E) = \infty$ for each $E \subset Z$ with $\mu(E) > 0$.*

*Remarks.* 1. This can be proved by the usual sort of transfinite argument but such arguments are not used at this stage in [**1**].
2. The general result follows: On $Y$, $\nu(\mathrm{E}) = \int_E f\,d\mu$ by Pl; on $Z$, $\nu(\mathrm{E}) = \int_E \infty\,d\mu$. The extension to $\sigma$-finite $\mu$ is standard.
3. The proof of the Lemma is very similar to that of [**1**, Chap. 11, Lemma 20].

*Proof.* In the following all sets called $E$, $E_i$ are assumed to be measurable and have finite $\nu$-measure.

If there exists no $E$ with $\mu(E) > 0$, take $Y = \varnothing$, $Z = X$; otherwise let $n_1$ be the smallest positive integer such that there exists $E_1$ with $\mu(E_1) > 1/n_1$.

If there exists no $E \subset X \setminus E_1$ with $\mu(E) > 0$ take $Y = E_1$, $Z = X \setminus E_1$; otherwise let $n_2$ be the smallest positive integer such that there exists $E_2 \subset X \setminus E_1$ with $\mu(E_2) > 1/n_2$.

We may assume that this process does not end; it yields $\{n_i\}$ nondecreasing with $\Sigma(1/n_i) < \Sigma\mu(E_i) \leqslant \mu(X) < \infty$, so $n_i \to \infty$. Let $Y = \bigcup E_i$, $Z = X \setminus Y$. The proof is concluded by showing that $\mu(E) = 0$ for all $E \subset Z$. (Remember that $\nu(E) < \infty$.) For all $m$, $E \subset X \setminus \bigcup\{E_i: 1 \leqslant i \leqslant m\}$; hence $\mu(E) \leqslant 1/(n_m - 1)$, so $\mu(E) = 0$.

REFERENCE

1. H. Royden, Real Analysis, 2nd. ed., MacMillan, New York, 1968.

# The Butterfly Curve

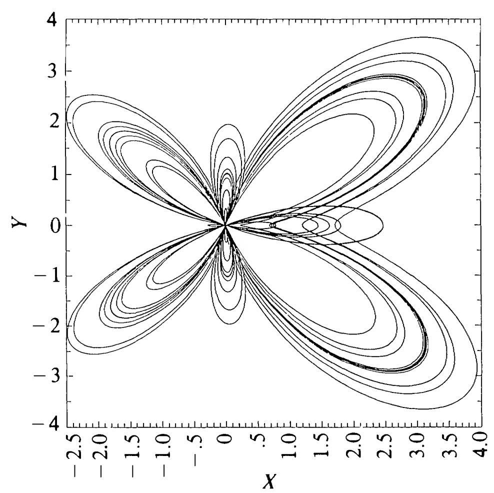
TEMPLE H. FAY
*Department of Mathematics, University of Southern Mississippi, Hattiesburg, MS 39406*

Many of us are attracted to the drawing of plane curves because of their inherent beauty expressed in symmetry, leaves and lobes, asymptotic behavior, asymmetry, or whatever is in the eye of the beholder. The classical rose curves and the like do not seem to provoke much student enthusiasm, but sketching more complicated curves, with the aid of a computer, seems to spark more interest and often brings surprises. There is something special about the dynamics of seeing a curve being drawn on the computer screen that makes the whole process more enjoyable.

It is the purpose of this note to point out that there is considerable pleasure and fun to be had with curve plotting while simultaneously obtaining teaching objectives. Pedagogically, the important questions of setting the bounds on the independent variable, the plotting step size, the screen window size, all need to be addressed prior to "hitting the RUN button." Here observations about size, period, symmetry, domain, and the like, take on new relevance for the student: minimize run time.

Moreover, there is an excitement generated by plotting curves whose shape is not a priori known. For example, the curves described by the polar equations $\rho = a + b\cos(n\theta)$ and $\rho^2 = a + b\cos(n\theta)$ are much more interesting than those described by $\rho = a\cos(n\theta)$ (see [**1**]). Students are often surprised by their behavior when $a < b$, $a > b$, $a = b$, $n$ is even, $n$ is odd, etc. Other interesting petal curves, for which students can predict the number of petals, are obtained from plotting the equations $\rho = (4\cos(n\theta) + \cos(m\theta))/\cos(\theta)$ with both $n$ and $m$ odd; if one of $n$ or $m$ is even, the curve has an asymptote and quite a different appearance (see [**2**]). For example, amusing curves are produced with $(n, m) \in \{(5,3),(3,5)(1,9),(3,2)\}$.

Perhaps the most interesting and beautiful of all the curves that the author has discovered is that of the "butterfly" displayed in the accompanying figure. The equation of the butterfly is $\rho = e^{\cos(\theta)} - 2\cos(4\theta) + \sin^5(\theta/12)$. The last term,




The Butterfly Curve: $\rho = e^{\cos(\theta)} - 2\cos(4\theta) + \sin^5(\theta/12)$

$\sin^5(\theta/12)$, is added purely to enhance the aesthetic appeal. Another butterfly-like curve has the equation $\rho = e^{\cos(2\theta)} - 1.5\cos(4\theta)$, which can be enhanced by adding other terms as above or by modifying the values of the constants.

REFERENCES

1. T. H. Fay and E. A. O'Neal, Counting the petals of rose curves, *Math. and Computer Ed.*, 17 (1983) 24–29.
2. T. H. Fay, Asymptotes in polar coordinates, *Math. and Computer Ed.*, 20 (1986) 19–31.

# Nonstandard Continuity and Uniform Convergence

CHRISTOPHER L. THOMPSON
*Faculty of Mathematical Studies, The University, Southampton S09 5NH, England*

We give a new nonstandard proof that the limit of a uniformly convergent sequence of continuous functions is continuous. Our proof is elementary in that it relies only on the transfer principle and makes no use of internal sets. As in many nonstandard proofs, the transitivity of $\approx$ (the relation of infinite closeness) is used to avoid an $\varepsilon/3$ argument.

THEOREM. *Let $f_n$ be continuous on $E \subseteq \mathbb{R}$ for $n = 1, 2, 3, \ldots$ and let $f_n$ converge uniformly to $g$ on $E$. Then $g$ is continuous on $E$.*

In the proof of the theorem we use the following nonstandard criterion for continuity. It is easily proved by transfer (exercise for the reader!) and is merely a slight variant of the usual nonstandard condition given, for example, by Robinson in [**4**].

LEMMA 1. *Let $f$ be a real-valued function on $E \subseteq \mathbb{R}$. Then $f$ is continuous at $c \in E$ if and only if ${}^*f(x) \approx f(c)$ for every $x$ in ${}^*E$ such that $x \approx c$ and $x - c$ is sufficiently small.*

We use the usual nonstandard criterion for uniform convergence (see [**4**]):

LEMMA 2. *A sequence $(f_n)$ converges uniformly to $g$ on $E$ if and only if ${}^*f_\nu(x) \approx {}^*g(x)$ for every $x$ in ${}^*E$ and every infinite $\nu$ in ${}^*\mathbb{N}$.*

*Proof of the theorem.* Suppose that $\nu$ in ${}^*\mathbb{N}$ is infinite, that $c \in E$, and that $x \in {}^*E$; then because $(f_n)$ converges uniformly to $g$ on $E$ we have

$$ {}^*g(x) \approx {}^*f_\nu(x) \tag{i}$$

$$ {}^*f_\nu(c) \approx g(c). \tag{ii}$$

Since each $f_n$ is continuous at $c$ we may write the usual $\varepsilon, \delta$ continuity criterion, apply transfer, and then take $\varepsilon$ infinitesimal. We obtain:

$$ {}^*f_\nu(x) \approx {}^*f_\nu(c) \tag{iii}$$

provided that $x \approx c, x \in {}^*E$, and $x - c$ is sufficiently small (depending on $\nu$). Combining (i), (ii), and (iii) we obtain

$$ {}^*g(x) \approx {}^*f_\nu(x) \approx {}^*f_\nu(c) \approx g(c), \tag{iv}$$

provided that $x \approx c$, $x \in {}^*E$, and $x - c$ is sufficiently small. By Lemma 1 it follows that $g$ is continuous at $c$ and the theorem is proved.

The proofs of this theorem in the textbooks on nonstandard analysis that I have consulted are more elaborate [**2**], more advanced [**1**], or else are very close in spirit to the conventional $\varepsilon/3$ argument [**3**, **4**].

We conclude with an observation about the validity of (iv). We have just seen that (iv) holds if $\nu$ is infinite and $x - c$ is a sufficiently small infinitesimal. Suppose now that $x \approx c$ and $x \in {}^*E$; then ${}^*g(x) \approx g(c)$ because $g$ is continuous, by the usual nonstandard criterion for continuity. Combining this with (i) and (ii) we see that (iv) is in fact true if $\nu$ is infinite, $x \in {}^*E$, and $x - c$ is infinitesimal, and without any further restriction.

REFERENCES

1. M. Davis, Applied Nonstandard Analysis, Wiley, New York, 1977.
2. J. M. Henle and E. M. Kleinberg, Infinitesimal Calculus, MIT Press, Cambridge, 1979.
3. A. E. Hurd and P. A. Loeb, An Introduction to Nonstandard Real Analysis, Academic Press, New York, 1985.
4. A. Robinson, Non-Standard Analysis, North-Holland, Amsterdam, 1974.

EDITOR'S NOTE

The paper by Oscar Cámpoli in the November, 1988, issue of the MONTHLY contained an elementary proof of the fact that $\mathbb{Z}[\frac{1}{2} + \frac{1}{2}\sqrt{-19}\,]$ is a principal ideal domain that is not a Euclidean domain. However, equally simple and more general proofs had already appeared elsewhere; see "Note on non-Euclidean principal ideal domains" by Kenneth S. Williams (*Math. Mag.*, 48 (1975) 176–177) and "A principal ideal ring that is not a Euclidean ring" by J. C. Wilson (*Math. Mag.*, 46 (1973) 34–38).

## SOLUTIONS OF ELEMENTARY PROBLEMS

### Distances with Specified Multiplicities

E 2938* [1982, 273]. *Proposed by P. Erdös, Hungarian Academy of Sciences.*

Can one find $n$ points in the plane (no three on a line, no four on a circle) so that for every $i$, $i = 1, 2, \ldots, n-1$, there is a distance determined by these points that occurs exactly $i$ times?

*Editorial Comment.* It is known that such configurations exist for $n = 3, 4, 5, 6, 7, 8$. Professor Erdös writes, "I am sure that for $n > n_0$ there is no solution, and I offer fifty dollars ($50) for the first proof of that assertion. On the other hand, I offer five hundred dollars ($500) for examples with arbitrarily large values of $n$."

We give here examples of the desired configurations for $n = 3, 4, 5, 6, 7, 8$.

For $n = 3$ consider the points $(0,1)$, $(0,0)$, and $(1,0)$; the distance $\sqrt{2}$ occurs once and the distance 1 occurs twice.

For $n = 4$ consider the points $(-1,0)$, $(0,0)$, $(1,0)$, and $(0,\sqrt{3})$; the distances $\sqrt{3}$, 1, and 2 occur once, twice, and three times respectively.

For $n = 5$ consider the points $(0,0)$, $(6,0)$, $(3,3\sqrt{3})$, $(3,\sqrt{3})$, and $(-2\sqrt{6}, 2\sqrt{3})$. The distances $2\sqrt{18+6\sqrt{6}}$, $2\sqrt{9+3\sqrt{6}}$, $2\sqrt{3}$, and 6 occur one, two, three, and four times respectively. This example is due to Carl Pomerance.

For $n = 6$ consider the points $(0,0)$, $(6,0)$, $(3,3\sqrt{3})$, $(3,\sqrt{3})$, $(-2\sqrt{6},2\sqrt{3})$, and $(3-\sqrt{6}, -\sqrt{3}-3\sqrt{2})$. The distances $2\sqrt{15+6\sqrt{6}}$, $2\sqrt{18+6\sqrt{6}}$, $2\sqrt{3}$, $2\sqrt{9+3\sqrt{6}}$, and 6 occur one, two, three, four, and five times, respectively.

For $n = 7$ consider the points $(0,0)$, $(2,0)$, $(2+\sqrt{6},\sqrt{2})$, $(\sqrt{6}+1,\sqrt{2}+\sqrt{3})$, $(1,\sqrt{3}+2\sqrt{2})$, $(0,2\sqrt{2})$, and $(1,2\sqrt{2}-\sqrt{3})$. The distances $2\sqrt{5-\sqrt{6}}$, $2\sqrt{3-\sqrt{6}}$, $2\sqrt{2}$, 2, $2\sqrt{3}$, and $2\sqrt{3+\sqrt{6}}$ occur one, two, three, four, five, and six times respectively. This example is due to Ilona Palásti, "On the seven points problem of P. Erdös," *Studia Scientiarum Mathematicarum Hungarica*, 22 (1987), 447–448.

For $n = 8$ consider the points $(0,2)$, $(2\sqrt{3},0)$, $(4\sqrt{3},0)$, $(5\sqrt{3},5)$, $(3\sqrt{3},9)$, $(\sqrt{3},7)$, $(3\sqrt{3},7)$, and $(2\sqrt{3},4)$. The distances 2, $2\sqrt{19}$, $2\sqrt{21}$, $2\sqrt{3}$, 4, $2\sqrt{7}$, and $2\sqrt{13}$ occur one, two, three, four, five, six, and seven times respectively. This example occurs in the paper by Ilona Palásti, "Lattice-point examples for a question of Erdös," to appear in *Periodica Mathematica Hungarica*.

### A Real Inequality

E 3227 [1987, 787]. *Proposed by Bradley Lucier, Purdue University, West Lafayette, IN.*

Assume that $\lambda > 0$, $0 < \Delta < y$, and $m > 2$. (Here $m$ is not necessarily an integer.) Show that

$$\Big\{(y+\Delta)^{1/(m-1)} + \lambda\big((y+2\Delta)^{m/(m-1)} - 2(y+\Delta)^{m/(m-1)} + y^{m/(m-1)}\big)\Big\}^{m-1}$$
$$-\Big\{y^{1/(m-1)} + \lambda\big((y+\Delta)^{m/(m-1)} - 2y^{m/(m-1)} + (y-\Delta)^{m/(m-1)}\big)\Big\}^{m-1}$$
$$\leqslant \Delta.$$

(This problem arose in studying the regularity of numerical approximations to the so-called porous medium equation:

$$\partial_t u(x,t) - \partial_x^2 (u(x,t))^m = 0.)$$

*Solution by the proposer*. With parameters as specified above, we must show $F(y+\Delta) - F(y) \leqslant \Delta$, where

$$F(y) = \left[y^{1/(m-1)} + \lambda\left[(y+\Delta)^{m/(m-1)} - 2y^{m/(m-1)} + (y-\Delta)^{m/(m-1)}\right]\right]^{m-1}.$$

By the Mean Value Theorem, it suffices to show that $F'(y) \leqslant 1$ for $y \geqslant \Delta$.

Let $\alpha = m/(m-1)$ and $f(y) = y^\alpha$. For $0 < \Delta < y$, we may expand $f$ by Taylor series to get

$$f(y+\Delta) - 2f(y) + f(y-\Delta) = 2\sum_{k=1}^{\infty}\binom{\alpha}{2k} y^{(1/(m-1))-(2k-1)}\Delta^{2k}.$$

Substituting this expression into $F(y)$ yields

$$F(y) = y\left(1 + \sum_{k=1}^{\infty} b_{m,k} y^{1-2k}\right)^{m-1},$$

where $b_{m,k} = 2\lambda\Delta^{2k}\binom{\alpha}{2k}$. For $y > \Delta$, the series converges. Note that the coefficients $b_{m,k}$ are all positive, because $1 < \alpha < 2$.

Differentiating, we obtain

$$F'(y) = \left(1 + \sum_{k=1}^{\infty} b_{m,k} y^{1-2k}\right)^{m-2}\left[1 + (m-1)\sum_{k=1}^{\infty} b_{m,k}\left(\frac{1}{m-1} + 1 - 2k\right)y^{1-2k}\right].$$

Since $(m-1)[1/(m-1) + 1 - 2k] \leqslant (m-1)[1/(m-1) - 1] = -(m-2)$ and the $b_{m,k}$ are all positive, we find

$$\begin{aligned}F'(y) &\leqslant \left(1 + \sum_{k=1}^{\infty} b_{m,k} y^{1-2k}\right)^{m-2}\left[1 - (m-2)\sum_{k=1}^{\infty} b_{m,k} y^{1-2k}\right]\\ &= (1+a)^{m-2}(1-(m-2)a),\end{aligned}$$

where $0 < a = \sum_{k=1}^{\infty} b_{m,k} y^{1-2k}$.

Setting $g(a) = (1+a)^{m-2}\{1-(m-2)a\}$, we compute $g'(a) = (m-2) \times (1+a)^{m-3}(1-m)a$. Since $m > 2$, we have $g'(a) < 0$ for $a > 0$. Thus $F'(y) \leqslant g(a) < g(0) = 1$, which completes the proof.

The only other solution received was incorrect.

### Sufficient Conditions for the Zero Matrix

E 3228 [1987, 787]. *Proposed by David K. Cohoon, Temple University, Philadelphia, PA.*

Let $S$ be an $m$ by $m$ matrix over $\mathbb{C}$. It is well known that $S^2 = S$ and trace($S$) $= 0$ imply that $S$ is the zero matrix. For which positive integers $n > 2$ and $m > 1$ does the pair of conditions $S^n = S$ and trace($S$) $= 0$ imply that $S$ is the zero matrix?

*Solution by John Henry Steelman, Indiana University of Pennsylvania, Indiana, PA.* The implication holds if and only if $m < p$, where $p$ is the smallest prime divisor of $n-1$.

Suppose $S^n = S$ and $\operatorname{tr}(S) = 0$. Because $S$ satisfies $x^n - x = 0$, the minimal polynomial for $S$ must divide $x^n - x$; hence each non-zero eigenvalue of $S$ is an $(n-1)$th root of unity. Let $\omega$ be a primitive $(n-1)$th root of unity, and suppose that $\omega^k$ has multiplicity $c_k$ as an eigenvalue of $S$, for $0 \leqslant k \leqslant n-2$. Let $k_1, \dots, k_r$ be the values of $k$ such that $\omega^k$ is an eigenvalue of $S$; note that $r \leqslant m$. Because $\operatorname{tr}(S) = 0$, $\omega$ is a root of $h(x) = \sum_{i=1}^{r} c_{k_i} x^{k_i}$, a polynomial with integer coefficients. Since any polynomial over $\mathbb{Z}$ satisfied by a primitive $(n-1)$th root of unity is satisfied by all primitive $(n-1)$th roots of unity, we have $h(\omega^q) = 0$ for all integers $q$ relatively prime to $n-1$.

In particular, suppose $m$ is less than the smallest prime divisor $p$ of $n-1$. Then $h(\omega^q) = 0$ for $1 \leqslant q \leqslant r$. This leads to a system of equations

$$(c_{k_1}, \dots, c_{k_r}) \begin{bmatrix} \omega^{k_1} & \omega^{2k_1} & \cdots & \omega^{rk_1} \\ \omega^{k_2} & \omega^{2k_2} & \cdots & \omega^{rk_2} \\ \vdots & \vdots & & \vdots \\ \omega^{k_r} & \omega^{2k_r} & \cdots & \omega^{rk_r} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

of Vandermonde type. Because $\omega \neq 0$ and the $\omega^{k_i}$'s are distinct, the system is non-singular, forcing $c_{k_1} = \cdots = c_{k_r} = 0$. Therefore, all eigenvalues of $S$ are 0, so that $S$ is nilpotent; thus $S^l = 0$ for sufficiently large $l$. Since $S^{n^t} = S$ for all positive integers $t$, we conclude that $S = 0$.

On the other hand, suppose $m \geqslant p$. Let $S_p$ be a $p$ by $p$ permutation matrix of a $p$-cycle and let

$$S = \begin{bmatrix} S_p & 0 \\ 0 & 0 \end{bmatrix}.$$

Then $\operatorname{tr}(S) = 0$ and $S^n = S$, but $S \neq 0$.

Note that we have proved the following: Counting multiplicities, the smallest number of $(n-1)$th roots of unity summing to zero is the smallest prime dividing $(n-1)$.

*Editorial comment.* Most solvers used, explicitly or implicitly, the result just stated on sums of roots of unity. O. P. Lossers attributed the result to Schoenberg and supplied the reference H. B. Mann, "On linear relations between roots of unity," *Mathematika* 12 (1965), 107–117. Note that since $S_p$ is a real matrix, the conclusion of the problem is still valid when restricted to real matrices.

Also solved by B. Dickinson, J. Delany, M. R. Darafsheh (Iran), J. Ferrer (Spain), S. M. Gagola, R. Gilmer, C. Khare (England), C.-K. Li, O. P. Lossers (The Netherlands), H. Schmidt, and the proposer. One incorrect and three incomplete solutions were received.

### A Concurrence in the Eternal Triangle

E 3231 [1987, 876]. *Proposed by Herbert Guelicher, Muenster, West Germany.*

In a triangle $P_1P_2P_3$, let $p_i$ be the side opposite vertex $P_i$, and let $s_i$ be a line parallel to $p_i$ (but different from $p_i$). Suppose that $s_i$ divides $P_iP_{i+1}$ in the (signed)

ratio $\lambda_i$, so that if $s_i$ meets $p_{i-1}$ in $Q_i$, then $\lambda_i = P_iQ_i/Q_iP_{i+1}$. (Subscripts are taken modulo 3.) Prove that the lines $s_1, s_2, s_3$ are concurrent if and only if

$$\lambda_1\lambda_2\lambda_3 - (\lambda_1 + \lambda_2 + \lambda_3) = 2.$$

*Editors' summary of solutions.* Many solvers obtained the required condition by a direct application of similar triangles. Most of the others introduced an appropriate coordinate system—rectangular, barycentric, or an equivalent vectorial representation—and obtained the result algebraically.

A typical solution using rectangular coordinates runs as follows. Note that since affine transformations preserve lines and ratios of lengths of parallel segments, we may assume without loss of generality that $P_1 = (0,0)$, $P_2 = (1,0)$, $P_3 = (0,1)$ in the standard coordinate system. Then $s_1$ is the locus of points $(x, y)$ with $x + y = \lambda_1(1+\lambda_1)^{-1}$, $s_2$ is the locus with $x = (1+\lambda_2)^{-1}$, and $s_3$ is the locus with $y = (1+\lambda_3)^{-1}$. We have concurrency of the $s_i$ if and only if

$$\lambda_1(1+\lambda_1)^{-1} = (1+\lambda_2)^{-1} + (1+\lambda_3)^{-1},$$

which is equivalent to the required condition.

A typical solution using barycentric coordinates is similar to the preceding. We assign to each point barycentric coordinates $(a_1, a_2, a_3)$, $\Sigma a_i = 1$, with respect to $P_1, P_2, P_3$. We then see that the line $s_i$ is the locus of points $(a_1, a_2, a_3)$ with $a_i$ equal to $t_i = (1+\lambda_i)^{-1}$, and that the three lines are concurrent if and only if $\Sigma t_i = 1$, which is equivalent to the desired relation.

Some contributors gave equivalent formulations based on signed areas. It was pointed out by E. Lee that if $\Delta P_1P_2P_3$ is of area 1 and $T_1, T_2, T_3$ have barycentric representations $T_i = (a_{i1}, a_{i2}, a_{i3})$ with respect to $P_1, P_2, P_3$, then the signed area of $\Delta T_1T_2T_3$ is $\det(a_{ij})$. Now, if $T_i$ is the intersection of $s_{i-1}$ and $s_{i+1}$, we have $T_1 = (t_1 + t, t_2, t_3)$, $T_2 = (t_1, t_2 + t, t_3)$, $T_3 = (t_1, t_2, t_3 + t)$, where $t_i = (1+\lambda_i)^{-1}$ and $t = 1 - \Sigma t_i$. By adding to the first column of $\det(a_{ij})$ the sum of the other columns, then subtracting the first row from the other rows, we find that the area of $\Delta T_1T_2T_3$ is

$$\det(T_1, T_2, T_3) = t^2 = \left(1 - \Sigma(1+\lambda_i)^{-1}\right)^2.$$

It follows that the lines are concurrent if and only if $\Sigma(1+\lambda_i)^{-1} = 1$.

This generalizes to higher dimensions as follows. Let $P_1P_2 \cdots P_n$ be a simplex of volume 1 in $R^{n-1}$. Let $s_1, \ldots, s_n$ be hyperplanes, with the $i$th barycentric coordinate of $s_i$ equal to $t_i = (1+\lambda_i)^{-1}$, where $\lambda_i$ has the same meaning as in the original problem, namely it is the ratio into which $s_i$ divides each of the edges $P_iP_j$. Then these hyperplanes bound a simplex with vertices $T_1, \ldots, T_n$, where the $j$th barycentric coordinate of $T_i$ is $t_j$ for $j \neq i$ and is $t_i + t$ for $j = i$, where $t = 1 - \Sigma t_i$. The volume of the simplex $T_1T_2 \cdots T_n$ is given by the determinant with rows $T_1, \ldots, T_n$. Making the column and row operations as before, we find that this volume is $t^{n-1} = (1 - \Sigma(1+\lambda_i)^{-1})^{n-1}$. Thus the given hyperplanes are concurrent if and only if $\Sigma(1+\lambda_i)^{-1} = 1$.

The volume formula in terms of barycentric coordinates used in the preceding paragraph can be deduced from the well-known determinant formula in terms of cartesian coordinates by applying a volume-preserving affine transformation which sends the simplex $P_1P_2 \cdots P_n$ to standard position with one vertex at the origin

and the adjacent edges along the axes. (Cf. H. S. M. Coxeter, *Introduction to Geometry*, New York, Wiley, 1961, page 219, where the case $n = 3$ is given.)

The proposer also provided a proof of the generalization to higher dimensions. M. S. Klamkin and A. Liu obtained a solution for the original problem as a consequence of a theorem in their preprint "Simultaneous generalization of the theorems of Ceva and Menelaus." J. E. Wetzel also provided a short solution of the problem and pointed out that if one uses the ratios $\mu_i = P_iQ_i/P_iP_{i+1}$ rather than $\lambda_i = P_iQ_i/Q_iP_{i+1}$, then the concurrency condition takes the simpler form $\mu_1 + \mu_2 + \mu_3 = 2$. In this formulation, the original requirement that the line $s_i$ be distinct from the edge $p_i$ can be removed.

Solved by J. C. Binz (Switzerland), H. Demir (Turkey), J. Dou (Spain), R. Dybvik (Norway), J. Ferrer (Spain), J. Fukuta (Japan), H. Guggenheimer, J. Heuver, W. Janous (Austria), H. Kappus (Switzerland), L. Kuipers (Switzerland), E. Lee, O. P. Lossers (The Netherlands), H. M. Marston, W. H. Oh (Korea), M. Orlowski & M. Pachter (South Africa), A. Pedersen (Denmark), S. Philipp, J. V. Savall (Spain), V. Schindler (East Germany), R. A. Simon (Chile), R. S. Tiberio, M. Vowe (Switzerland), and the proposer.

**One Reason Why 1987 Was Unique**

E 3238 [1987, 995]. *Proposed by Dieter Bode, Fachhochschule Lippe, Lemgo, West Germany.*

Show that 1987 is the only prime number $p$ satisfying

$$\pi(p) = q(p) \cdot \{\pi(q(p)) + \tau(q(p))\},$$

where $\pi(n)$ denotes the number of primes not exceeding $n$, $q(n)$ is the sum of the decimal digits of $n$, and $\tau(n)$ is the number of positive divisors of $n$.

*Solution by Lorraine L. Foster, Joseph Loo (student), and Yokichi Tanaka (student), California State University, Northridge, CA.* Define $f(n) = q(n)[\pi(q(n)) + \tau(q(n))]$. As verified by computer, the only solutions to $f(n) = \pi(n)$ with $n < 8467$ are $n = 415, 750, 870, 1696, 1987$. Among these, only 1987 is prime. We will show that $\pi(n) > f(n)$ for all $n > 8467 = n_0$ (this in fact holds for all $n > 1999$), and hence no other solutions exist (prime or otherwise).

Suppose first that $n_0 < n \leqslant n_1$, where $n_1 = 270{,}000$. By inspection, $q(n) \leqslant 46$, $\tau(q(n)) \leqslant 9$, and $\pi(q(n)) \leqslant 14$, so $f(n) \leqslant 1058$, and hence $\pi(n) \geqslant \pi(n_0) = 1059 > f(n)$.

Now suppose $n > n_1$. Note that if $q(n) < 9$, then $f(n) \leqslant 9(9 + 9) < \pi(n)$. Hence we may assume that $q(n) \geqslant 9$, so that $\pi(q(n)) < q(n)/2$, since $\pi(m) < m/2$ for $m > 8$. Also, if $q(n) \geqslant 9$, then $\tau(q(n)) \leqslant q(n)/2$, because $\tau(m) \leqslant m/2$ for $m > 8$. Thus $f(n) \leqslant (q(n))^2$.

The decimal representation of $n$ has at most $\log_{10} n + 1$ digits, so $q(n) \leqslant 9(\log_{10} n + 1) \leqslant 11 \log_{10} n$. The well-known density of primes yields $\pi(n) > (n \log 2)/(4 \log n)$ for $n \geqslant 2$. (cf. Niven and Zuckerman, *An Introduction to the Theory of Numbers*, Wiley (New York, 1960, 1966, 1972, 1980), §8.1.) Also, differentiation shows that $y = (\log x)^3/x$ decreases for $x > 20.1$. Combining these results for $n > n_1$, we have $(\log n)^3/n \leqslant (\log n_1)^3/n_1 < \log 2(\log 10)^2/(4 \cdot 11^2)$, and hence $f(n) \leqslant 11^2(\log n/\log 10)^2 < (n \log 2)/(4 \log n) < \pi(n)$, as claimed.

Also solved by Victor Pambuccian and the proposer.

### A Unique Function on Natural Numbers

E 3241 [1987, 996]. *Proposed by Gregory P. Wene, University of Texas, San Antonio.*

Suppose that $a, b, c$ are given natural numbers with $a < b < c$.

(i) show that a function $f\colon N \to N$ is uniquely defined by the following pair of formulas:

$$f(n) = n - a \qquad \text{if } n > c,$$
$$f(n) = f(f(n+b)) \quad \text{if } n \leqslant c.$$

(ii) Determine a necessary and sufficient condition for $f$ to have at least one fixed point.

(iii) Give such a fixed point explicitly in terms of $a$, $b$, and $c$.

*Solution by John Henry Steelman, Indiana University of Pennsylvania, Indiana, PA.* Suppose that $f$ satisfies the conditions above.

(i) We first show, by induction on $c - n$, that $f(n) = f(n + b - a)$ for all $n \leqslant c$. If $0 \leqslant c - n < b$, then $f(n) = f(f(n+b)) = f(n + b - a)$ because $n + b > c$. If $c - n \geqslant b$, then by induction we have $f(n) = f(f(n+b)) = f(f(n + 2b - a))$, since $n + b \leqslant c$. However, we also have $n + (b - a) \leqslant c$, so that $f(n + b - a) = f(f(n + 2b - a))$, which establishes the claim. When $n \leqslant c$, iterating this formula until the argument exceeds $c$ shows that $f$ is uniquely defined. Furthermore, the formula for $f(n)$ when $n \leqslant c$ is $f(n) = c + b - 2a - j$, where $j$ is the smallest non-negative residue of $c - n$ modulo $b - a$.

(ii) The function $f$ has a fixed point if and only if $b - a$ divides $a$. If $f$ fixes $n$, we must have $n \leqslant c$, in which case we can write $n$ as $n = c - k(b - a) - j$, where $k$ and $j$ are non-negative integers and $0 \leqslant j < b - a$. Using the formula for $f(n)$, we find that $f$ fixes $n$ if and only if $c - k(b-a) - j = n = f(n) = c + b - 2a - j$, which holds if and only if $(k+1)(b-a) = a$.

(iii) If $b - a$ divides $a$, the fixed points of $f$ consists precisely of those numbers having the form $c + b - 2a - j$, where $0 \leqslant j < b - a$.

Also solved by J. C. Binz (Switzerland), D. K. Brown & S. Phillip, D. Callan, P. Dennee (student), J. Ferrer (Spain), S. M. Gagola, W. Janous (Austria), O. P. Lossers (The Netherlands), A. Müller (Switzerland), M. D. Meyerson, A. Pedersen (Denmark), R. E. Prather, W. P. Wordlaw, Shippensburg Univ. Math. Problem Solving Group, and the proposer.

### When $k$-gonal Numbers are Fermat Numbers

E 3247 [1988, 51]. *Proposed by Nick MacKinnon, Winchester College, Winchester, England.*

Given a positive integer $k$ greater than 2, the $k$-gonal numbers are the terms of the sequence

$$1, k, 3k - 3, 6k - 8, \ldots,$$

which has constant second difference $k - 2$. The Fermat numbers are the terms of

the sequence

$$3 = 2^{2^0} + 1, \qquad 5 = 2^{2^1} + 1, \qquad 17 = 2^{2^2} + 1, \qquad 257 = 2^{2^3} + 1, \ldots.$$

Under what circumstances can a $k$-gonal number be a Fermat number?

*Solution by Stephen M. Gagola, Jr., Kent State University, Kent, OH.* For $k > 2$, this cannot happen after the second $k$-gonal number. The $n$th $k$-gonal number is $(k-2)n(n-1)/2 + n$. Suppose

$$\frac{n(n-1)}{2}(k-2) + n = 1 + 2^{2^m} \tag{$*$}$$

with $n, k > 2$. The equation $(*)$ shows that $n - 1$ cannot be divisible by any odd prime, so $n - 1 = 2^r$ is a power of 2. Since $n > 2$, we have $r \geqslant 1$, so $n$ is odd and $2^r \equiv -1 \pmod n$. The equation $(*)$ also implies that $n$ is a factor of $1 + 2^{2^m}$, so $2^{2^m} \equiv -1 \pmod n$. The second congruence shows that the order of the integer 2 in the group of units mod $n$ is exactly $2^{m+1}$, so by the first congruence $r$ must be an odd multiple of $2^m$. Thus $n = 1 + 2^r \geqslant 1 + 2^{2^m}$. Since $n$ is a factor of $1 + 2^{2^m}$, we conclude $n = 1 + 2^{2^m}$ and $(k-2)n(n-1)/2 = 0$, contrary to $k > 2$.

Also solved by J. G. G. Cuartas (Columbia), N. Franceschine, T. Hermann (Hungary), W. Janous (Austria), H.-S. Ki (student, Korea), K.-W. Lau (Hong Kong), H. E. Lomeli (student, Mexico), O. P. Lossers (The Netherlands), K. McInturff, J.-M. Monier (France), R. E. Prather, J. T. Ward, W. P. Wardlaw, and the proposer. Two incorrect solutions were received.

**Partial Sums of Partial Fractions**

E 3250 [1988, 131]. *Proposed by Lennart Bondesson, Umea, Sweden.*

Suppose $N > 1$ and $a_1 < a_2 < \cdots < a_N$ are given real numbers. Consider the partial fraction decomposition

$$\prod_{i=1}^{N} \frac{1}{a_i + x} = \sum_{i=1}^{N} \frac{A_i}{a_i + x}.$$

It is familiar that the numbers $A_1, A_2, \ldots$ alternate in sign. What can we say about the signs of the partial sums $A_1, A_1 + A_2, A_1 + A_2 + A_3, \ldots, A_1 + A_2 + \cdots + A_N$?

*Solution by the editors.* The partial sums also alternate in sign, beginning with a positive term, except that the last partial sum is zero. This follows by induction on $N$, using the identity

$$\prod_{i=1}^{N-1} \frac{1}{a_i + x} - \prod_{i=2}^{N} \frac{1}{a_i + x} = (a_N - a_1) \prod_{i=1}^{N} \frac{1}{a_i + x}. \tag{$*$}$$

Using the partial fraction expansions

$$\prod_{i=1}^{N-1} \frac{1}{a_i + x} = \sum_{i=1}^{N-1} \frac{B_i}{a_i + x} \quad \text{and} \quad \prod_{i=2}^{N} \frac{1}{a_i + x} = \sum_{i=2}^{N} \frac{C_i}{a_i + x}$$

of these shorter products, and equating corresponding coefficients in $(*)$, we obtain

$$\begin{aligned}
(a_N - a_1)A_1 &= B_1,\\
(a_N - a_1)(A_1 + A_2) &= B_1 + B_2 - C_2,\\
(a_N - a_1)(A_1 + A_2 + A_3) &= B_1 + B_2 + B_3 - C_2 - C_3,\\
&\vdots\\
(a_N - a_1)(A_1 + A_2 + \cdots + A_{N-1}) &= B_1 + B_2 + \cdots + B_{N-1} - C_2 - \cdots - C_{N-1},\\
(a_N - a_1)(A_1 + A_2 + \cdots + A_N) &= B_1 + B_2 + \cdots + B_{N-1} - C_2 - \cdots - C_N.
\end{aligned}$$

By induction

$$B_1 > 0, \qquad B_1 + B_2 < 0, \qquad B_1 + B_2 + B_3 > 0, \qquad \ldots$$

and

$$C_2 > 0, \qquad C_2 + C_3 < 0, \qquad C_2 + C_3 + C_4 > 0, \qquad \ldots,$$

so that

$$A_1 > 0, \qquad A_1 + A_2 < 0, \qquad A_1 + A_2 + A_3 > 0, \qquad \ldots.$$

More complicated solutions along similar lines were obtained by Ha-seo Ki and Kim McInturff. The proposer sketched a solution using complex analysis.

## ADVANCED PROBLEMS

6601. *Proposed by Sajal K. Das, University of North Texas, Denton, TX and the late Patrick O'Hara, University of Central Florida, Orlando.*

Given positive integers $q \geq 2$ and $k \geq 1$ let $\pi_k$ be the permutation of the set $\{0, 1, 2, \ldots, q^k - 1\}$ defined by

$$\begin{aligned}
\pi_k(i) &= qi - (q^k - 1)\lfloor qi/(q^k - 1)\rfloor && \text{if } i = 0, 1, 2, \ldots, q^k - 2,\\
\pi_k(i) &= i && \text{if } i = q^k - 1.
\end{aligned}$$

If $\pi_k$ is factored into disjoint cycles, determine the cycle lengths which occur and the number of cycles of each length.

6602. *Proposed by Allen R. Miller, Naval Research Laboratory, Washington, DC.*

Suppose $0 \leq k_1 \leq 1$, $0 \leq k_2 \leq 1$, and $\phi$ is real; if $k_1 = k_2 = 1$, suppose further that $\phi$ is not a multiple of $2\pi$. If

$$\begin{aligned}
\alpha &= e^{i\phi}(e^{i\phi} - k_1 k_2),\\
\beta &= k_1^2 + k_2^2 - 2k_1 k_2 e^{i\phi},\\
\gamma &= 2(k_1^2 + k_2^2 - k_1 k_2 \cos\phi - 1),
\end{aligned}$$

prove that the roots of the self-inversive quartic equation

$$\alpha z^4 + \beta z^3 + \gamma z^2 + \bar{\beta} z + \bar{\alpha} = 0$$

have absolute value one.

6603. *Proposed by Carl Pomerance, University of Georgia, Athens.*

Let $G_n$ denote the set of points in Euclidean $n$-space which can be reached from the origin in a finite number of steps of unit length, where each step lands on a point with rational coordinates. In other words $G_n$ is the additive subgroup of $Q^n$ generated by those points in $Q^n$ at distance one from the origin.

(a) Show that $G_n = Q^n$ if and only if $n \geq 5$.

(b) Obtain an arithmetic characterization of $G_1, G_2, G_3, G_4$.

## SOLUTIONS OF ADVANCED PROBLEMS

### Tale of a Series Involving the Tail of Another Series

6549 [1987, 559]. *Proposed by L. Matthew Christophe, Jr., Wilmington, DE.*

Sum the series

$$\sum_{k=1}^{\infty} (-1)^{k+1}\left(\frac{(2k)!}{2^{2k}(k!)^2 k(2k+1)}\right)\left(\ln 2 - \sum_{j=1}^{k} \frac{(-1)^{j+1}}{j}\right).$$

*Solution by L. E. Mattics, University of South Alabama, Mobile.* As we shall see this problem is closely related to problem E3140 [1986, 216; 1988, 57]. If $S$ is the value of the given sum, then

$$\begin{aligned} S &= -\sum_{k=1}^{\infty}\binom{2k}{k}\frac{1}{2^{2k}k(2k+1)}\int_0^1 \frac{x^k}{1+x}\,dx \\ &= \int_0^1 \frac{1}{1+x}\left(2\sum_{k=1}^{\infty}\binom{2k}{k}\frac{x^k}{2^{2k}(2k+1)} - \sum_{k=1}^{\infty}\binom{2k}{k}\frac{x^k}{2^{2k}k}\right)dx. \end{aligned}$$

We now rely on two formulas in D. H. Lehmer's article "Interesting Series Involving the Central Binomial Coefficient," this MONTHLY, 92 (1985) 449–457. Specifically, we replace $x$ by $x/4$ in Lehmer's formula (6) and we replace $x$ by $\sqrt{x}/2$ in the next to the last formula on Lehmer's page 451. Using these results, we obtain

$$\begin{aligned} S &= 2\int_0^1\left(\frac{\sin^{-1}\sqrt{x}}{\sqrt{x}(1+x)} - \frac{1}{1+x}\right)dx - \int_0^1 \frac{2\log\left\{2(1+\sqrt{1-x})^{-1}\right\}}{1+x}\,dx \\ &= 4\int_0^1 \frac{\sin^{-1}u}{(1+u^2)}\,du - 2\log 2 - 2(\log 2)^2 + 2\int_0^1 \frac{\log(1+\sqrt{1-x})}{1+x}\,dx. \end{aligned}$$

From problem E3140 we have

$$\int_0^1 \frac{\sin^{-1}u}{1+u^2}\,du = \frac{\pi^2}{8} - \int_0^1 \frac{\tan^{-1}u}{\sqrt{1-u^2}}\,du = \left(\log(\sqrt{2}+1)\right)^2/2$$

and also (with the aid of the substitution $x = 1/u^2$)

$$\int_0^1 \frac{\log(1+\sqrt{1-x})}{1+x}\,dx = 2\int_1^\infty \frac{\log(u+\sqrt{u^2-1})}{u(u^2+1)}\,du - 2\int_1^\infty \frac{\log u}{u(u^2+1)}\,du$$

$$= \left(\log(\sqrt{2}+1)\right)^2 + 2\int_0^1 \frac{v\log v}{v^2+1}\,dv$$

$$= \left(\log(\sqrt{2}+1)\right)^2 - \frac{\pi^2}{24}.$$

Consequently

$$S = 4\left(\log(\sqrt{2}+1)\right)^2 - \frac{\pi^2}{12} - 2\log 2 - 2(\log 2)^2.$$

*Editorial Comment.* The proposer informs us that this series arose in a problem of fractal geometry.

Solved also by C. Georghiou (Greece), I. E. Leonard (Canada), and Dr. Klaus Zacharias (Berlin). One incorrect solution was received.

# REVIEWS

EDITED BY JOSEPH KONHAUSER
*Department of Mathematics, Macalester College, St. Paul, MN 55105*

*Algorithms in Combinatorial Geometry.* By Herbert Edelsbrunner. Springer-Verlag, New York, 1987. xv + 423 pp.

JACOB E. GOODMAN
*Department of Mathematics, City College, City University of New York, New York, NY 10031*

Hand in hand with the computer revolution which has been taking the world by storm in the second half of this century, a related revolution has been taking place more quietly in our own back yard. In its more modest way it has also caused upheavals of a sort—shifts of student population masses, unemployment in mathematics departments, higher salaries for computer scientists, integration of computers into various areas of mathematical research, and increased attention by mathematicians to problems—often quite intriguing and difficult problems—arising in theoretical computer science, especially in the search for more and still more efficient algorithms to carry out the large-scale computations demanded by a host of applications, both theoretical and practical.

This offshoot of the computer revolution has made inroads into several branches of mathematics, among them combinatorics and geometry. It is with the latter that Edelsbrunner's new book is concerned: the field known as computational geometry.

What, exactly, is computational geometry? Generally speaking, it consists of the design and analysis of geometric algorithms; in practice, these usually deal with searching and sorting in the plane and in higher dimensions, encoding arrangements of geometric objects so that information can be retrieved efficiently, computing geometric objects associated to other geometric objects (such as a triangulation of a nonconvex polygon), and performing these procedures in provably optimal time and/or space. Typically, the dimension of the ambient space is low, often 2 or 3, while the size of the input is high.

Consider the following problem. You are given the coordinates of 10 points in the plane, and asked which of them are extreme points of the set, i.e., which of them do not lie in the convex hull of the rest. A trivial problem? Of course—just consider each point $P$ in turn, and sort the remaining points by their angles around $P$; if the angle between any two consecutive directions is greater than $180°$, $P$ is extreme.

But wait. What if there are 10,000 points instead of 10? Since sorting $n$ numbers takes about $n \log_2 n$ operations [**9**], and since this sorting must be done once around each of the points, it would take more than a billion operations to carry out the procedure with $n = 10{,}000$. But around 1970 Graham thought a little more about the geometry of the problem, and came up with an algorithm which was almost as simple [**4**] and which accomplished the same objective in time $O(n \log n)$. Since then several equally efficient algorithms have been found (the reader is invited to try his hand as well!). Thus 10,000 points can now be processed in a fraction of a second.

A word of caution is in order here, which applies, in fact, to theoretical computer science as a whole. When we say an algorithm has time complexity $O(f(n))$, we generally mean that it takes no more than $cf(n)$ operations to perform on an input

of size $n$, *for some constant* $c$. But the constant $c$ may be (and sometimes is, especially in more sophisticated algorithms) rather large. So what relevance can such theoretical results have to the practical implementations which are needed to turn algorithms into useful computer programs? There are two answers to this question. The more cynical is that computer scientists have made a sort of game of ignoring constants; they care only about the *asymptotic* performance of their algorithms, as $n \to \infty$; and since it never does in real life, such work is not necessarily of any practical significance. The second answer, which is a bit more oblique, is grounded in experience; it relates to what one might call the "unreasonable effectiveness of asymptotic complexity." It is that generally, as Tarjan has said **[15]**, good theoretical results often seem to lead to good practical results, a case in point being that Khachiyan's proof of the fact that linear programming is polynomial, with his theoretically ground-breaking but practically all but useless ellipsoid algorithm **[8]**, led eventually to the Karmarkar algorithm [7], which is turning out to be quite useful indeed. The reader is free to choose his own answer.

One is struck afresh, looking over the extensive bibliography of Edelsbrunner's book, at just how new a field computational geometry is. Precisely two pre-1970 papers are listed, each of them from 1962 and each one page long; there is a smattering of papers from the early 1970s, and it is not until the mid '70s that the full flood begins, following on the heels of Shamos' work (his 1978 thesis coined the phrase "computational geometry"). Of course these are the papers in *computational* geometry; in discrete geometry, on the other hand, specifically on topics relating to arrangements and polytopes, there are venerable papers going back to Steiner (1826), to Sylvester (1893), and to Dehn (1905), as well as more recent work by Coxeter, Erdös, Hadwiger, Grünbaum, and Klee among many others. (There are also some names which are perhaps more unexpected, such as Dirichlet, Carathéodory, Borsuk, or Milnor, whose 1964 results on the homology groups of real algebraic and semi-algebraic varieties [**12**] have played a key role in several recent papers.) This interaction between two fields, one well established and the other quite new, is at the core of Edelsbrunner's book.

In 1986 a summer research conference devoted to discrete and computational geometry took place. Its announcement spoke of fostering the partnership that had begun in recent years between geometry and computer science and of promoting an ongoing cross-fertilization between the two areas. Herbert Edelsbrunner, as much as any researcher in the field, embodies that goal. His book represents the first flowering of the collaboration that has been taking place between geometers and combinatorists on the one hand, and computer scientists on the other, a collaboration so lively as to blur even further the already tenuous distinction between mathematics and computer science. In addition to much material that appears for the first time, it touches on, among other results of recent years, the seminal paper of Lipton-Tarjan **[10]** which gave the first optimal-time and optimal-storage solution to the post office problem and the related point-location problem in $R^2$; Megiddo's linear-time algorithm for linear programming in fixed dimension $d$ [**11**]; Ungar's beautiful solution by circular sequences of P. R. Scott's problem on the maximum number of directions determined by $n$ noncollinear points **[16]**; Seidel's convex hull algorithm in dimension $d$ [**14**], which is optimal for $d$ even; the Goodman-Pollack generalization of sorting to higher dimensions **[3]**, in which the $\lambda$-matrix encoding of a point configuration was introduced; the optimal algorithm of Edelsbrunner-

O'Rourke-Seidel [2] for finding and representing compactly the cell complex associated to an arrangement of hyperplanes, with its many applications to everything from detecting degeneracies and finding minimal-volume simplices in point configurations to computing $\lambda$-matrices and constructing higher-order Voronoi diagrams; the Hart-Sharir improvement of Szemerédi's bound on the length of Davenport-Schinzel sequences [5], and their connection with the upper envelopes of functions considered by Pach-Sharir [13]; and the Clarkson probabilistic method [1] and the related $\varepsilon$-net method of Haussler-Welzl [6], which together are responsible for initiating one of the hottest areas of current research in computational geometry and which are yielding surprising combinatorial results as well.

Edelsbrunner's book begins with the premise that arrangements of hyperplanes, and their dual objects, configurations of points, are at the heart of computational geometry, and its first third is devoted to the study of such arrangements, discussing along the way such things as the Borsuk-Ulam theorem in topology, the Dehn-Sommerville relations on the number of faces of each dimension of a polytope, Helly's theorem on convex sets, the "ham sandwich" dissection theorem, the "upper bound theorem" for polytopes, Voronoi diagrams, and the "circular sequence" and "order-type" encoding of arrangements. Its second section explores a host of basic geometric algorithms based on the material in the first part, including algorithms for encoding arrangements, for finding convex hulls in all dimensions, for linear programming in low dimensions, and for planar point-location. Finally, there are several chapters on applications of the preceding material to such things as visibility graphs, minimum spanning trees, and the post office problem. It concludes with a remarkable chapter in which the problem of finding line-transversals for a given family of line segments is first analyzed, geometrically and combinatorially, and then several basic techniques for designing algorithms are applied successively, the result being a set of algorithmic solutions to several variants of the original problem. Anyone wishing to enter the field of computational geometry would be well-advised to work carefully through this chapter.

The book has a number of commendable features, not least among them the numerous graded problems, ranging from simple exercises (rated "1") to unsolved research problems ("5"), relating both to the text and to relevant work in the literature (and woe to the author of a paper who finds his hard-earned results relegated to a level of "1" or "2"). Also quite useful are the historical remarks to be found at the end of each chapter, which trace the development of the subject from its classical period (i.e., the 1970s and sometimes even earlier!) to the present.

Computational geometry is indeed a new and vigorous field, a field dominated by prolific young people. To be sure, the other side of the energy which the young workers in computational geometry have brought to the field has been an apparent willingness to rush into print with more and more refinements of algorithms not yet published even in conference proceedings, and to compete with one another to knock a $\log n$ or a $\log\log n$ off last month's time bound. Through this effusion of instantly obsolete results has flowed, however, a steady stream of first-rate work (is this very different, one wonders, from the situation in mathematics as a whole?), much of which has found its way into the present volume.

When a book is as comprehensive as the one at hand, it is perhaps inevitable that a few errors should creep in. The only serious one I could find occurs on page 118, in the midst of a discussion of the Euler relation for a convex polytope, when the

reader is told that René Descartes had a proof of the relation before Ludwig Euler did. Surely the author must mean *Blanche*?

For a fascinating glimpse into this new area of research, which has burgeoned so much during its brief existence that a new journal has been already been created in its honor, or for an in-depth study of some of the major work going on at the moment in discrete mathematics and theoretical computer science, I heartily recommend *Algorithms in Combinatorial Geometry*.

REFERENCES

1. K. L. Clarkson, New applications of random sampling in computational geometry, *Discrete Comput. Geom.*, 2 (1987) 195–222.
2. H. Edelsbrunner, J. O'Rourke, and R. Seidel, Constructing arrangements of lines and hyperplanes with applications, *SIAM J. Comput.*, 15 (1986) 341–363.
3. J. E. Goodman and R. Pollack, Multidimensional sorting, *SIAM J. Comput.*, 12 (1983) 484–507.
4. R. L. Graham, An efficient algorithm for determining the convex hull of a finite planar set, *Inform. Process. Lett.*, 1 (1972) 132–133.
5. S. Hart and M. Sharir, Nonlinearity of Davenport-Schinzel sequences and of a generalized path-compression scheme, *Combinatorica*, 6 (1986) 151–177.
6. D. Haussler and E. Welzl, $\varepsilon$-nets and simplex range queries, *Discrete Comput. Geom.*, 2 (1987) 127–151.
7. N. Karmarkar, A new polynomial-time algorithm for linear programming, *Combinatorica*, 4 (1984) 373–395.
8. L. G. Khachiyan, Polynomial algorithm for linear programming, *Dokl. Akad. Nauk SSSR*, 244 (1979) 1093–1096.
9. D. E. Knuth, Sorting and Searching: The Art of Computer Programming III, Addison-Wesley, Reading, MA, 1973.
10. R. J. Lipton and R. E. Tarjan, Applications of a planar separator theorem, in Proc. 18th Ann. IEEE Sympos. Found. Comput. Sci., 1977, pp. 162–170.
11. N. Megiddo, Linear programming in linear time when the dimension is fixed, *J. Assoc. Comput. Mach.*, 31 (1984) 114–127.
12. J. Milnor, On the Betti numbers of real varieties, *Proc. Amer. Math. Soc.*, 15 (1964) 275–280.
13. J. Pach and M. Sharir, The upper envelope of piecewise linear functions and the boundary of a region enclosed by convex plates: combinatorial analysis, *Discrete Comput. Geom.* (to appear).
14. R. Seidel, A convex hull algorithm optimal for point sets in even dimensions, Rep. 81–14, Dept. Comput. Sci., Univ. British Columbia, Vancouver, BC.
15. R. E. Tarjan, Mathematics in computer science, invited address at Amer. Math. Soc. Centennial, Providence, RI, 1988.
16. P. Ungar, 2N noncollinear points determine at least 2N directions, *J. Combin. Theory Ser. A*, 33 (1982) 343–347.

## *Elementary Number Theory*. By Charles Vanden Eynden, Random House, 1987. xii + 266 pp. *Elementary Number Theory and its Applications*, 2nd ed. By Kenneth H. Rosen. Addison-Wesley, 1988, xiii + 466 pp.

EMIL GROSSWALD
*Department of Mathematics, Temple University, Philadelphia, PA 19122*

What is number theory? The Pythagoreans considered it almost as a religion; Gauss called it the queen of mathematics; *Time* magazine (as quoted by A. Weil in his book *Number Theory*) describes it as "an abstruse specialty concerned with the properties of whole numbers." And there are many other descriptions offered for number theory. There is also a diversity of views concerning its usefulness. On the one hand, Hardy claimed that what he was doing (and much of it was number

theory) was perfectly useless—and that is the way he wanted it. He even added the stronger statement that, if he could be shown some use for his mathematics, he would change the subject of his studies. On the other hand, many authors go out of their way to convince the reader that number theory is very useful and has numerous practical applications. Some authors even add that these uses alone justify the study of the discipline.

At present it seems that at least in this respect a consensus has developed: those who study number theory do it because of the beauty of the subject matter and their own fascination with its problems, without any thought concerning the usefulness of number theory, and then it turns out (surprisingly?) that the results obtained are eminently applicable to "real-life problems," hence useful.

In spite of this diversity of opinions and even if the practitioners of number theory are unable to define it precisely, they know quite well if a given problem, or method, belongs to number theory, or not.

This is perhaps less trivial than it may seem, if we take into account the great diversity of problems that are considered and methods that are used in number theory. We have here for instance work like that of Hardy, Littlewood and others, which uses such a heavy analytic machinery in the study of number theoretic problems that A. Weil considers the work not "analytic number theory," as it is usually called, but simply analysis. It is not obvious what this "analytic number theory" has in common with, say, sieve theory, or probabilistic number theory, or the theory of ideals, etc. Nevertheless, all these are branches of number theory and no number theorist will doubt it.

It is a fact that throughout history most great mathematicians have worked, at least occasionally, on number theoretic problems. The reason seems to be that one cannot help being fascinated by number theory. And why is this the case? Here, presumably, the opinions may differ, but at least one reason seems to be the way in which the simplicity of the questions asked is in contrast with the sophistication of the needed proofs. To give an example, consider the following simple question: Is it true, or false, that every even integer larger than 4 is the sum of two odd primes? Up to this day the answer (suspected to be "yes") is not known and the proof of the best known approximation to it, namely, that for every integer $n$, either $2n = p + q$, or $2n = p + qr$ (with $p, q, r$ primes) requires very difficult consideration of sieve theory.

In spite of all the diversity of topics, methods, applicability, etc. that exist among the (by now numerous) branches of number theory, all are built upon the same foundation, the so-called "elementary number theory," which treats some standard topics needed in the more specialized branches of number theory. Topics belonging to elementary number theory are, among others, the nature of integers, divisibility, prime numbers, arithmetical functions, congruences, and, perhaps, primitive roots, quadratic residues and quadratic reciprocity. The number of textbooks of elementary number theory is quite large. They all treat the stated topics and most of them add some enrichment, that is, somewhat less elementary topics, selected according to the idiosyncrasies of their authors. Some of these textbooks are so much alike that one cannot help wondering whether all of them were really needed.

This cannot be said about the two books reviewed here. To start with, they differ very substantially from each other in style and both of them, each in its own way, are really excellent. Although the books are quite different, it is noteworthy that neither ignores computers and that both devote considerable space to cryptology. A

sign of our time, it seems. Neither of the books contains a proof of the prime number theorem, which is only natural in an elementary text.

The book by Vanden Eynden is written so that it should appeal even to the least talented of students. It may well be suitable for a variety of audiences, but then, for the more sophisticated reader, there are many other textbooks available that are either more condensed, or cover more material. However, for the unsophisticated reader, either for use in the classroom, or for self-study, I know of no better book. It is totally informal; every topic is carefully motivated and described before a definition is given; the proofs are first explained on hand of numerical examples, and in the proofs themselves every i is dotted and every t is crossed. The induction principle is used consistently and is almost the unifying theme for most proofs.

There are numerous exercises, graded as A, B, or C and those in category C sometimes require considerable ingenuity.

Of real interest are also the nice biographies of mathematicians and the historical notes.

It is natural that these virtues come at a price. The number of topics covered is more limited than that of the book by Rosen and some important theorems (e.g., the theorems of Fermat, Euler, and Wilson ) are discussed rather briefly. Also, some very interesting topics are relegated to the exercises. Furthermore, in view of the excellent presentation of the public-key cryptography and the RSA method, the reader (at least this reviewer) may regret the absence of a discussion of the knapsack method.

The book by Rosen has "applications" in its title and, indeed, it presents an impressive number of applications. It is kept much more formal than Vanden Eynden's. In fact, the first chapter almost suggests that the treatment of number theory will be pursued in a strictly axiomatic manner. It is to be hoped that this will not scare away any reader, because, in fact, in spite of its (relative) formality the book is, pedagogically speaking, quite sound. It is meant for a more sophisticated readership than the book by Vanden Eynden and may be more appropriate for use in a classroom than for self-study. However, it also has leisurely introductions to proofs and numerous instances in which particular cases of theorems are first worked out on numerical examples. Still, the motivation is more limited than in Vanden Eynden's book and some definitions precede the discussion of the topic.

The amount of material covered is very large. Not only are Fibonacci numbers, a perpetual calendar and round-robin tournaments treated, but also such topics as pseudoprimes, tests for primality, Carmichael numbers, the splicing of telephone cables—and even Kaprekar's constant! This is made possible by the larger number of pages and by the somewhat more condensed presentation. A whole chapter is devoted to cryptology and both the knapsack method and the RSA method are discussed in detail. Also, at the end of each paragraph one finds, after the usual exercises, a number of "Computer Projects," which are suggestions for writing computer programs. Some of the exercises, as well as some computer programs, are quite difficult.

Both books end with arithmetical tables (the ones in Vanden Eynden are more extensive) and have ample and very useful bibliographies; it may be worthwhile to mention that each of the bibliographies lists the name of the author of the other book.

Finally, one should say a word about the material make-up of the books. Both are well printed and attractive. Both books have their share of typographical errors

and occasionally these can be quite confusing for a naive reader. However, this reviewer knows from personal experience how difficult it is to avoid these errors even after the most painstaking, repeated proofreading and shall make no further comments about them.

*Combinatorics of Finite Sets.* By Ian Anderson. Oxford University Press, New York, 1987. xv + 250 pp.

Daniel J. Kleitman
*Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139*

This volume describes a branch of mathematics that is dear to my heart, and here is why.

When I was growing up, physics was a glamorous subject. Atomic weapons had affected the course of history, terminating World War II with spectacular abruptness. Potential applications and employment opportunities were promising, and the intellectual challenge of learning what physical nature was all about was a stimulus to studying it.

Mathematics on the other hand, though my favorite subject in school, seemed to present limited opportunities. The only careers plausible in mathematics appeared to be academic, and at the time almost all academic institutions discriminated to some degree against Jews. Aiming at a career in mathematics seemed to be seeking out trouble.

These thoughts and the fact that my older brother chose to major in physics led me to make the same choice.

As an undergraduate I studied as much mathematics as physics, and in the process became vaguely aware that I had some unusual abilities in that area. Pierre Samuel, who was visiting from France, gave a double graduate course in Bourbaki-like mathematics, one of the courses I had signed up to take. He gave several assignments that required ingenuity for their solution. While walking one day I suddenly 'saw' the way to solve one of these. Later this same 'seeing' happened again, and several times more on his final examination.

I did not find the same experience in physics.

In graduate school I studied physics, and, though I had no particular feel for the subject, I was able to complete my degree in physics and do work in that field for eight years.

In 1961 I met a former classmate, David Lubell, who showed me an idea of his which he hoped to apply to a problem he had found in a book by Stan Ulam. I found his idea beautiful and stimulating, and thought about his problem. That night I 'saw' a solution. I borrowed the book, and in a few days found the solution to another problem in it.

These problems had been given to Ulam by Paul Erdös. For a few years I was not able to catch up with Erdös, but when he did hear of my work, he sent me more problems of the same general flavor, and I was fortunate enough to be able to solve several of these as well.

At this point Erdös wrote, "Why are you only a physicist? Why not be a mathematician?" I informed him that I would be happy to be a mathematician. As a result of his efforts I was offered a position in mathematics at MIT, where I remain today.

The subject matter of *Combinatorics of Finite Sets* is that of Lubell's idea and of all those Erdös problems.

The field has a number of virtues: its problems are easy to state and to understand without any background, and are challenging because there are no standard methods of attack on them.

The methods that can be applied have a beauty to them and the problems and methods have a generality which makes them actually applicable, upon occasion, to analysis of algorithms in computer science, for example, and elsewhere in mathematics as well.

Sperner, in 1929, addressed the following question. Suppose $C$ is a collection of subsets of an $n$-element set $S$ such that no element of $C$ contains another. How large can $C$ be? He gave an inductive proof that $C$ could have no more elements than the number of subsets of $S$ of cardinality $[n/2]$.

Here are two quick proofs of this result (one of which was Lubell's idea) which illustrate some of these points.

Let a chain be a collection of subsets of $S$ such that every pair of its elements has one containing the other. We can prove Sperner's result if we can partition the subsets of $S$ into chains each one of which has an element of cardinality $[n/2]$ since at most one element of $C$ can be in each chain.

Let the elements of $S$ be the natural numbers up to $n$. Represent each subset $T$ by a sequence of length $n$ whose $j$th member is ')' if $j$ is in $T$ and '(' otherwise. In any such sequence, close any adjacent ( ) pair, and any such pairs separated by already closed pairs. Now put two subsets in the same chain if their closed parts are identical and in the same positions. It is straightforward to verify that this defines a partition into chains each one of which contains a subset of cardinality $[n/2]$. (End of first proof.)

Consider all $n!$ maximal chains. Each chain intersects $C$ at most once; thus the proportion of the chains that contain a cardinality $k$ element of $C$, when summed over $k$, is at most 1. Since every cardinality $k$ set occurs in the same number of these chains as any other, the proportion of the cardinality $k$ sets that are in $C$ is the proportion of the chains that contain such sets. It follows that the proportion of the cardinality $k$ sets in $C$, summed over $k$, is at most 1. Sperner's theorem is an immediate consequence of this inequality. (End of Lubell proof.)

Another question in this area is: suppose we want to choose $k$ subsets of an $n$-element set $S$, each having cardinality $j$, so that the fewest cardinality $j-1$ subsets are contained in one or more of them. How can we do it?

The answer: order the elements of $S$ from 1 to $n$, and describe each subset of $S$ by a sequence of 0's and 1's, making the $r$th entry of a set $A$ a 1 when the $r$th element is in $S$. Then choose the subsets of cardinality $j$ whose sequences are the smallest when interpreted as numbers.

This statement has a beautiful and simple proof. If instead we try to choose $k$ such subsets to minimize the number of cardinality $j$ subsets whose sequences differ in two places from at least one of the chosen $k$, the answer in general is unknown.

Anderson's volume provides a lucid discussion of many of the basic results in this area, with many challenging exercises as well. This is my favorite subject. I owe my career to a chance encounter with it. I consider it the ideal introduction to mathematical reasoning for students at any level and anyone else for that matter. I find the subject more fun than crossword puzzles or chess, and vastly more rewarding. I think you should buy and read this book.

*For All Practical Purposes: An Introduction to Contemporary Mathematics.* By COMAP. Project Director: Solomon Garfunkel. Coordinating Editor, Lynn A. Steen. W. H. Freeman and Company, New York, 1988. xii + 450 pp. + 26 video tapes and Course Guide. ISBN 0-7167-1830-8.

ARTHUR GITTLEMAN
*Computer Science and Engineering Department, California State University, Long Beach, CA 90840*

What should students who only take one mathematics course study? Most available texts take one of two approaches both of which I find inappropriate. The first is a rehash of elementary mathematics. The typical treatments of sets, integers, rationals, geometry, and similar topics are dull. Mathematics appears dead and useless. The sole virtue of these texts (from a publisher's point of view) is their dual use in liberal arts mathematics courses and for elementary school teacher preparation.

The second type of text isn't dull, only useless. It has pictures and puzzles to entertain. One can do tricks with numbers and shapes. Such a text also has virtues. Students and instructors survive each other. Many professors find liberal arts mathematics difficult to teach and are happy at least to have a text which quells resistance.

When I sought a text for the course, I wanted neither the "spinach is good for you" nor the "candy keeps them quiet" approaches. I wanted good taste and nutrition. Something palatable and substantial. The *Monthly* Telegraphic Review of *For All Practical Purposes* intrigued me. This was just what I had in mind for the course; real mathematics really being used. A college graduate would appreciate the role of mathematics in our society. He or she would be able to participate more intelligently in the decision making of a technically oriented democracy.

When I received the text, I was pleasantly surprised to learn that 26 half-hour videotapes were available. The Consortium for Mathematics and Its Applications (COMAP) received a grant to produce a telecourse in mathematics and another grant to produce the textbook. The authors suggest that the book may be used alone or in concert with the television shows. For students enrolled in a telecourse a Course Guide provides an overview of each program, skill objectives, and a sample shorts-answer examination. The authors "hope that these materials will provide a rich and exciting environment in which to learn more about the power and centrality of mathematics in our world."

Their environment is rich, exciting, and very bumpy. The television shows come the closest to meeting the COMAP objectives. The textbook would be difficult to use by itself and is not a satisfactory complement to the shows. The telecourse guide is helpful but needs much improvement. If you can handle the rough ride, the materials are usable now. The shows with a revised text and telecourse guide would be ideal.

The materials cover five major areas: Management Science, Statistics, Social Choice, On Size and Shape, and Computers. Most areas are divided into four parts. The book has twenty-one chapters, one for each part. The 26 television shows include one for each part and one overview show for each area.

I taught liberal arts mathematics this past summer (1988) using the text, telecourse guide, and videotapes. In the 45-hour course I covered three Management

Science chapters, four from the Social Choice area and two from On Size and Shape. We viewed eleven half-hour shows. The topics are meaty and require thorough chewing. I chose to omit the units on Statistics and Computers because their treatment in the book (I didn't view these shows) was relatively mundane and because students may take other courses in these areas.

The thrust of *For All Practical Purposes* is excellent. It depicts mathematics in economically, socially, politically, and scientifically relevant contexts. It introduces deep mathematical algorithms. The television shows are professionally produced. They motivate the student but do not overwhelm. The amiable host injects much humor. The overview shows excellently introduce each area.

The management science overview shows the need for mathematics in scheduling the vast array of component tasks for the Apollo launch. Analyzing the tasks involved in unloading and loading a passenger plane shows the scheduling problem in a simpler example. They pose questions such as whether speeding up one task will reduce the overall time taken. Viewers clearly get the point that the answers to these questions are not intuitive and require mathematical analysis.

Shows on individual topics expand upon the examples in the overview show. They model the problem mathematically and begin to develop algorithms for solutions. These shows are at the right level. The examples, which are realistic rather than contrived, convince students of the need for mathematics. Students cannot master the algorithms just by watching the shows but they are not "snowed." After seeing a show, they are attentive to a fuller presentation of the mathematics.

Better-written, supporting materials combined with the shows would make an ideal course meeting COMAP and my objectives. The textbook is interesting but too difficult and disjointed. I still have not been able to comprehend some of this text. Box 1.6 on page 16 is an aside attempting to show good eulerizations for rectangular networks. Three figures are shown and the last paragraph of the box purports to explain how the figures were drawn. I need a clearer explanation.

The fascinating Chapter 13 classifies patterns. Figure 13.7 shows the seven different one-dimensional strip patterns. Because the figure was taken from hand-painted pottery it was difficult to see whether or not a symmetry was present. The 17 two-dimensional patterns were shown in Figure 13.8, labelled intriguingly, but with no explanation in the text. They introduce the two Penrose tiles that they say can cover the plane nonperiodically but not periodically. Both their explanation and Figure 13.4 are incomprehensible. I was fascinated enough to seek other references. They score 10 out of 10 on introducing fascinating topics, but only a 1 on the clarity and completeness of the explanations.

Organization within chapters is poor. Chapter 9 on social choice confused me. It presents various voting methods while at the same time comparing results under sincere and insincere voting. Presenting the voting methods first would much improve this chapter. The first example of a mixed strategy game in Chapter 11 is more complicated than the second. In Chapter 12, the method they give for fair division between two persons does not generalize to fair division among more than two persons. Yet they refer with no explanation to an example of fair division among four players in Table 12.1.

The exercises at the end of each chapter are not graded and are often too difficult. Exercise 4 of Chapter 2 asks for the conditions on $m$ and $n$ that guarantee an $m$ by $n$ grid graph has a Hamiltonian circuit. This is too abstract for students

taking just one mathematics class. It should have been placed at the end of the 28 chapter exercises.

I would not recommend using the textbook without the videotapes. I suggest especially careful preparation if the book is used with the tapes. The instructor needs to omit the confusing sections and must always provide much additional explanation. I selected several algorithms from each chapter to concentrate on and tried to show students how to solve selected problems. By limiting what they were expected to master I was able to bring the material within their level.

Because the textbook was deficient I also used the telecourse guide. A short chapter in this guide corresponds to each chapter in the text. The multiple-choice questions and the long answer problems are at the right level. They give the student needed practice. Several answers are incorrect however. The later chapters are inadequate. The chapter on patterns has no long answer problems on patterns. The superficiality of the guide material on measurement contrasts with the difficulty of the textbook problems. I could not use either set of problems and omitted those chapters. The telecourse guide is not enough to support the shows. The earlier chapters are satisfactory but the later ones are poorly done. More text would be needed for this guide to be used as intended.

The television shows are by far the best part of the package. My only criticism concerns the dramas. They often use little skits to make a point. I enjoyed the skits but some students thought they were corny. Excessively stupid people finally saw that their naive approaches did not work. Mathematics not only solved the problem but also got rid of these inane people. I accept this dramatic absurdity because it is pedogogically effective. What I objected to were the stereotypes. For example, in dramatizing voting procedures they had the rural yokels pitted against the city slickers.

I like the television shows and look forward to using them again. They bring out the students' creativity. I respect these students. Some are poor at mathematics but others have as much mathematical ability as science and mathematics majors. They find mathematics irrelevant when treated as a puzzle or even an art. They have a humanistic feeling for the integration of art with life and respond to the combination of algorithms and applications presented in the tapes. Students completing this course should be able to speak precisely about the importance of mathematics. They need to learn some algorithms to make their understanding more than empty words. We must be careful not to obscure the richness of vigorous applications with too many mathematical details. The television shows give just the right balance.

The next time I teach the course I may prepare my own written explanations and problems to go along with the tapes. I hope there will be revisions of the text and telecourse guide. Students will thrive and grow and that is what general education in mathematics should achieve.

# TELEGRAPHIC REVIEWS

*Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.*

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

| | | |
|---|---|---|
| T: Textbook | P: Professional Reading | 1–4: Semesters |
| C: Computer Software | L: Undergraduate Library | **: Special Emphasis |
| S: Supplementary Reading | 13: Grade Level | ??: Questionable |

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

**General, S, C.** *Challenging Mathematical Problems with PASCAL Solutions: A Sourcebook of Problems for Computers.* Donald D. Spencer. Camelot, 1988, 144 pp, $13.95 (P). [ISBN: 0-89218-086-2] The first section contains elementary problems from algebra, geometry, probability, statistics, mathematical recreations, etc., with cartoon illustrations; a second section contains a PASCAL program solution for each problem. JNC

**Mathematics Appreciation, S(12-14), L.** *Visual Patterns in Pascal's Triangle.* Dale Seymour. Dale Seymour Publications, 1986, v + 138 pp, (P). [ISBN: 0-86651-304-3] Numerous variants of Pascal's triangle (congruences, multiples, hidden patterns) illustrated in visual displays presented in form suitable for overhead transparency. A marvelous opportunity for student discovery of relations between arithmetic and geometric patterns. LAS

**Mathematics Appreciation.** *Mathematical Maxims and Minims.* Nicholas J. Rose. Rome Pr, 1988, vii + 169 pp, $13.95 (P). A delightful collection of anecdotes, quotes, cartoons, and poems all in celebration of mathematics in all its forms, from Abstract to Zero. SB

**Mathematics Appreciation, L.** *Heureka!* Heinrich Hemme. Vandenhoeck & Ruprecht, 1988, 109 pp, DM 19,80 (P). [ISBN: 3-525-40732-7] A collection of 95 mathematical puzzles requiring insight but little knowledge. Includes at least one solution for each, and information about sources. JD-B

**Precalculus, T(13: 1).** *Precalculus.* Margaret L. Lial, Charles D. Miller. Scott Foresman, 1989, 680 pp, $28. [ISBN: 0-673-15872-1 Following a review of algebra, this text focuses on the study of functions including trigonometric, exponential, and logarithmic; also includes chapters on systems of equations and inequalities and analytic geometry. Also available: ***Instructor's Guide and Solutions Manual, Instructor's Answer Manual,*** and ***Student's Solutions Manual and Study Guide.*** JNC

**Precalculus, T(13: 2).** *College Algebra and Trigonometry.* Adelbert F. Hackert, Gene R. Sellers. Harcourt Brace Jovanovich, 1988, x + 787 pp, $29. [ISBN: 0-15-507920-4] Four chapters containing high school algebra topics are followed by coverage of functions (including exponential, logarithmic, trigonometric, and inverse functions), and chapters on conic sections, systems of equations and inequalities, the binomial theorem, and mathematical induction. A *Student Solutions Manual* and an *Instructor's Manual* are available. JNC

**Education, P, L.** *The Effects of Modern Mathematics.* Sándor Klein. Akademiai Kiado, 1987, 436 pp, $48. [ISBN: 963-05-4023-1] Detailed report and analysis of four "new math" elementary school programs inspired by the work of Zoltán P. Dienes implemented in the 1960's and 1970's in specific schools in four locations (Budapest, New York, Quebec, and Brazil). Eighteen specific hypotheses and questions reflecting goals of the "new math" programs were investigated by careful statistically controlled methods. General conclusion: programs in Budapest and Brazil supported all hypotheses of the study; those in Quebec and New York were not clear. Evaluations on which this comparative study was based were completed over ten years ago. LAS

**History, P, L**.** *Mathematical Visions: The Pursuit of Geometry in Victorian England.* Joan L. Richards. Academic Pr, 1988, xiii + 266 pp, $34.95. [ISBN: 0-12-587445-6] Scholarly account of how new geometrical theories of the last half of the nineteenth century were received and interpreted in England. Examines the position of mathematics in English higher education and its role in the development of English mathematics. Traces the impact of culture and sociology and educational philosophy on the shaping of mathematics in Victorian England. Unlike other historical accounts, there is little effort to get the general reader to comprehend details of the subject matter. Not easy reading but worth the while. Copiously footnoted. Over seventeen pages

of references including many to works of, for example, Cayley, Clifford, DeMorgan, Klein, Russell, and Whewell. Epilogue on the demise of "English descriptive geometry" with illuminating discussion of the Tripos examinations. JK

**Logic, P.** *Mathematical Intuitionism: Introduction to Proof Theory.* A.G. Dragalin. Transl. of Math. Mono., V. 67. AMS, 1988, ix + 228 pp, $75. [ISBN: 0-8218-4520-9] Presents important methods of proof theory in intuitionistic logic. Covers syntactical methods, intuitionistic arithmetic, algebraic models, freely growing sequences, cut elimination in higher-order intuitionistic logic. Assumes background of introductory logic course. Extensive bibliography. Translation of 1979 Russian edition. KS

**Logic, T(14: 1), S*, L.** *Introduction to Proofs in Mathematics.* James Franklin, Albert Daoud. Prentice-Hall, 1988, vii + 175 pp, $16.95 (P). [ISBN: 0-7248-1009-9] An example-rich introduction to mathematical proofs. Begins by *proving* specific arithmetic properties instead of confirming them by calculation. Notes following each example explain, in detail, motivation of technique and encourages reader to generalize. Nice explanation of predicates and quantifiers, negation, and conditionals first without, then with symbolic language. Emphasis on intuition, not truth tables. Examples drawn from calculus, linear algebra, and number theory. Also includes sections on sets, induction, vector spaces, limits, counting, functions and applications to computer programs. Each section ends with a number of exercises (graded easy to difficult) with selected solutions in the back. SB

**Logic, P.** *A Formalization of Set Theory without Variables.* Alfred Tarski, Steven Givant. AMS Colloq. Pub., V. 41. AMS, 1987, xxi + 318 pp, $60. [ISBN: 0-8218-1041-3] This monograph presents new logical formalisms within which set theory and number theory can be developed. Language uses no variables, quantifiers, or connectives; instead, statements are equations between expressions involving four operation symbols and symbols for membership and identity relations. KS

**Foundations, L.** *The Reality of Numbers: A Physicalist's Philosophy of Mathematics.* John Bigelow. Clarendon Pr, 1988, viii + 193 pp, $45. [ISBN: 0-19-824957-8] Written from the perspective of metaphysical materialism, this provocative book challenges the notion that mathematical objects can be defined into existence. The engaging text is suffused with examples that are almost tactile. MU

**Number Theory, S(18), P.** *Divisors.* Richard R. Hall, Gérald Tenenbaum. Tracts in Math., V. 90. Cambridge U Pr, 1988, xvi + 167 pp, $39.50. [ISBN: 0-521-34056-X] Classical analytic treatment of the probabilistic distribution of divisors of natural numbers. Includes many recent results on sequences of asymptotic density one. Each chapter concludes with exercises suitable for a diligent graduate student. GG

**Group Theory, S(17-18), L.** *Existentially Closed Groups.* Graham Higman, Elizabeth Scott. Math. Soc. Mono., V. 3. Clarendon Pr, 1988, xiii + 156 pp, $49.95. [ISBN: 0-19-853543-0] A group $M$ is said to be existentially closed if every finite set of equations and inequalities which is soluble over $M$ is actually soluble in $M$. While this theory was initially part of the theory of infinite groups, it has also become part of model theory and recursion theory. This introductory book does not present all that is known of such groups; instead, the book's intent is to give the reader enough understanding and inspiration to pursue the subject further. The text is well written and accessible to anyone with some basic knowledge of group theory and logic. LW

**Algebra, S(18), P.** *Linear Algebraic Monoids.* Mohan S. Putcha. London Math. Soc. Lect. Note Ser., V. 133. Cambridge U Pr, 1988, x + 171 pp, $24.95 (P). [ISBN: 0-521-35809-4] Written to stimulate research interest in the area of linear algebraic monoids, the author presents a largely self-contained and comprehensive treatment of the current state of the theory, resting largely on work done by the author and Lex Renner. Accessible to graduate students; extensive bibliography. JS

**Algebra, P.** *Derivatives, Nuclei and Dimensions on the Frame of Torsion Theories.* Jonathan S. Golan, Harold Simmons. Pitman Res. Notes in Math. Ser., V. 188. Longman Scientific & Technical (US Distr: Wiley), 1988, 120 pp, $44.95 (P). [ISBN: 0-582-03448-5] The category of unitary left modules over a ring can tell you a lot about the ring. How do you study this category? By considering the "frame" (a certain type of lattice) of all so-called torsion theories over this category? This frame is studied in this monograph. SG

**Algebra, T(15-16), S, L.** *Sets and Groups: A First Course in Algebra.* J.A. Green. Routledge & Kegan Paul, 1988, xi + 258 pp, $16.50 (P). [ISBN: 0-7102-1227-5] Revised edition of book first published in 1965. Includes discussion of rings, fields, and vector spaces as well. New edition writes maps on left. Many examples and exercises (selected solutions at end). Small type is somewhat difficult to read, but overall a nice introduction to abstract algebra. SB

**Algebra, P.** *Brauer Groups and the Cohomology of Graded Rings.* Stefaan Caenepeel, Freddy van Oystaeyen. Pure & Appl. Math., V. 121. Marcel Dekker, 1988, x + 261 pp, $89.75. [ISBN: 0-8247-7978-9] A study of Azumaya algebra over $Z$-graded commutative rings including graded Galois theory, projective graded modules, graded cohomology groups, and graded Brauer groups. SG

**Calculus, T(13-14: 1, 2).** *Calculus with Analytic Geometry, Second Edition.* Dennis G. Zill. PWS-Kent, 1988, xiv + 1042 pp, $39.50. [ISBN: 0-534-91620-1] A standard comprehensive calculus book which features particularly nice graphics and biographical sketches. This edition includes more exercises, examples, figures, and applications. CEC

**Calculus, S**(13).** *Schaum's 3000 Solved Problems in Calculus.* Elliott Mendelson. McGraw-Hill, 1988, v + 442 pp, $19.95 (P). [ISBN: 0-07-041480-

7] Excellent source for extra examples, homework, or test problems in calculus sequence. Begins with important precalculus material; inequalities, absolute value, graphs of functions and circles. No text, but some concepts are generated by problems. Most problems are standard exercises, but this book contains many more problems than any textbook, e.g., 61 curve sketching, 58 volume, 116 infinite series, 115 power series. SB

**Calculus, S(13-14).** *Die mathematischen Abenteuer von Fritz und Katharina.* Klaus Langmann. Vandenhoeck & Ruprecht, 1988, 141 pp, DM 19,80 (P). [ISBN: 3-525-40733-5] A collection of 77 stories introducing the sort of problems in calculus, differential equations, and linear algebra encountered in the early semesters by college students of mathematics, science, and engineering. Sketches of solutions. JD-B

**Complex Analysis, T(18: 2-4), S**, P**, L.** *The Logarithmic Integral, I.* Paul Koosis. Stud. in Adv. Math., V. 12. Cambridge U Pr, 1988, xvi + 606 pp, $89.50. [ISBN: 0-521-30906-9] An inviting, enthusiastic, readable, comprehensive introduction to the logarithmic integral and its many roles in real analysis, complex analysis, and functional analysis. Expository style is informal, intuitive, and self-contained—treatment should be accessible to anyone with graduate training in analysis. Includes several dozen problems, often with copious hints. A mine of beautiful mathematics, "done in public." PZ

**Differential Equations, P.** *Fuchsian Differential Equations With Special Emphasis on the Gauss-Schwarz Theory.* Masaaki Yoshida. Aspects of Math., V. E11. Friedr. Vieweg & Sohn, 1987, xiv + 215 pp, (P). [ISBN: 3-528-08971-7] Studies linear ordinary differential equations and systems of partial differential equations with finite dimensional solution spaces in the complex analytic category. Concentrates on Gauss-Schwarz theory of the hypergeometric equation and its higher-dimensional generalization. SP

**Partial Differential Equations, P.** *Lecture Notes in Mathematics-1357: Partial Differential Equations and Calculus of Variations.* Ed: S. Hildebrandt, R. Leis. Springer-Verlag, 1988, vi + 423 pp, $37.10 (P). [ISBN: 0-387-50508-3] A collection of eighteen papers, some of them surveys, by associates of SFB 72 (Sonderforschungsbereich 72), a research institute at the University of Bonn. Emphasis is on existence and regularity results, nonlinear differential equations, special equations of mathematical physics, and problems of scattering theory. SP

**Partial Differential Equations, T(18: 2), P.** *Diffusion Processes and Partial Differential Equations.* Kazuaki Taira. Academic Pr, 1988, xviii + 452 pp, $69.50. [ISBN: 0-12-682220-4] Studies the relationship between probability and partial differential equations. Contains the necessary background material from real and functional analysis, manifolds, Sobolev spaces. Emphasis is the study of degenerate elliptic differential operators of second order followed by the existence and construction of Markov processes in terms of boundary value problems. GN

**Operator Theory, S(18), P.** *Direct and Inverse Scattering on the Line.* Richard Beals, Percy Deift, Carlos Tomei. Math. Surveys & Mono., No. 28. AMS, 1988, xiii + 209 pp, $53. [ISBN: 0-8218-1530-X] Linear ordinary differential operators of arbitrary order are studied by constructing sets of special eigenfunctions (generalized Jost solutions). The inverse problem—to reconstruct the operator from special eigenfunction data—is also considered. Applications to spectral theory, quantum-scattering theory, and nonlinear wave mechanics. SP

**Operator Theory, S(18), P.** *Introduction to Operator Theory and Invariant Subspaces.* Bernard Beauzamy. Math. Lib., V. 42. North-Holland (US Distr: Elsevier Science), 1988, xiv + 358 pp, $84.25. [ISBN: 0-444-70521-X] Assuming some background in operator theory and functional analysis, the presentation is in six parts: general theory, compactness, Banach algebra techniques, analytic functions, dilation and extensions, and invariant subspaces. Research areas encountered along the way are pointed out and open questions are raised. Bibliography, index, and some interesting commentary. JS

**Operator Theory, P.** *Surveys of Some Recent Results in Operator Theory, Volume I.* Ed: John B. Conway, Bernard B. Morrel. Pitman Res. Notes in Math. Ser., V. 171. Longman Scientific & Technical (US Distr: Wiley), 1988, 259 pp, $59.95 (P). [ISBN: 0-582-00519-1] Collection of papers presented during special year in operator theory at Indiana University. Contains recent results concerning Bergman and Bloch spaces, interpolation problems, spectral decompositions, spectral sets, and invariant subspaces. ES

**Functional Analysis, P.** *Lecture Notes in Mathematics-1317: Geometric Aspects of Functional Analysis.* Ed: J. Lindenstrauss, V.D. Milman. Springer-Verlag, 1988, vii + 289 pp, $25.80 (P). [ISBN: 0-387-19353-7] Third volume of Proceedings from Israel Seminar 1986-87 session. Includes articles concerning several relations between finite dimensional convexity theory and Banach space theory, Read's invariant subspace results, Banach lattices, spectral theory, inequalities for volumes, and norms of convex bodies, zonoids, and more. ES

**Analysis, T*(15-16: 1), L*.** *Introduction to Applied Mathematics.* L. Sirovich. Texts in Appl. Math., V. 1. Springer-Verlag, 1988, xii + 370 pp, $39.95. [ISBN: 0-387-96884-9] First in a new Springer series of texts in applied mathematics designed to reinforce traditional methods and to meet current and future needs of advances in areas such as numerical and symbolic computer systems, dynamical systems, and chaos. Based on notes from a course at Brown University for juniors and seniors majoring in applied mathematics, engineering, and other sciences. Complex analysis (enough to serve as a preliminary to a traditional course), discrete linear systems, Fourier series and applications, orthogonal polynomials, Fourier and Laplace transforms,

and partial differential equations. Informal, student-friendly, and readable. Stress on building student confidence. Examples; exercises without answers; of length manageable in one semester; prerequisites are linear algebra and ordinary differential equations. JK

**Analysis, T(17: 1), S.** *Problems in Distributions and Partial Differential Equations.* C. Zuily. Math. Stud., V. 143. North-Holland (US Distr: Elsevier Science), 1988, 245 pp, $79 (P). [ISBN: 0-444-70248-2] Presents the theory of distributions through sets of problems with complete solutions. The first six chapters cover the classical theory, the seventh chapter discusses applications to partial differential equations. Assumes some knowledge of $L^p$-spaces. Good text to use for an independent study of the subject. GN

**Geometry, T(16-17: 1), S, P, L.** *Projective Geometry.* Pierre Samuel. Transl: Silvio Levy. Undergrad. Texts in Math. Springer-Verlag, 1988, x + 156 pp, $35 (P). [ISBN: 0-387-96752-4] Developed for a course for prospective teachers at the ENSET in France. A concise, lucid exposition of projective spaces over fields of characteristic $p$. JNC

**Algebraic Topology, S(18), P, L.** *Cyclic Homology of Algebras.* Peter Seibt. World Scientific, 1987, xi + 160 pp, $20 (P); $37. [ISBN: 9971-50-470-7; 9971-50-468-5] Based on lectures given in 1985 at the University of Marseille Luminy, this exposition of cyclic cohomology is not meant as an introduction to the theory but rather as an attempt to identify the basic algebraic facts and techniques of the theory. The first chapter is concerned with the connections between cyclic (co)homology and Hochschild (co)homology, while the second considers the relations to deRham theory and Lie theory in the case of characteristic zero. At the end of each chapter are some remarks to guide the interested reader to the appropriate papers accompanied by a seemingly complete list of references.

**Algebraic Topology, P, L.** *Geometric and Algebraic Topology.* Ed: Henryk Tourńczyk, Stefan Jackowski, Stanisław Spież. Banach Center Pub., V. 18. PWN, 1986, 417 pp. [ISBN: 83-01-07321-7] This volume consists of 26 articles collected from the topology semester held at the International Stefan Banach Mathematical Center in spring 1984. Expository in nature, the articles are arranged into sections entitled: Low-dimensional manifolds; Higher-dimensional manifolds; Group actions; Differential topology and geometry; Cyclic homology and homology of groups; Shape theory and its homological aspects; Dimension theory and theory of continua. LW

**Topology, P.** *Braids.* Ed: Joan S. Birman, Anatoly Libgober. Contemp. Math., V. 78. AMS, 1988, xxxv + 730 pp, $62 (P). [ISBN: 0-8218-5088-1] Proceedings of a 1986 research conference containing thirty papers on algebraic, geometric, and physical aspects of the theory of braids. SG

**Dynamical Systems, P.** *Lecture Notes in Mathematics-1347: Iterates of Piecewise Monotone Mappings on an Interval.* Chris Preston. Springer-Verlag, 1988, 166 pp, $16.30 (P). [ISBN: 0-387-50329-3] Notes surveying the dynamics of continuous piecewise monotone maps of the interval into itself. The central concern is the asymptotic behaviour of iterates of a point in the interval, with an emphasis on properties of such maps which are invariant under topological conjugacy—here a continuous monotone change of coordinates. Results of Guckenheimer, Misiurewicz, Milnor-Thurston, Parry, as well as the author are presented. The treatment is self-contained and accessible from a background in basic real analysis and the topology of the real line. CE

**Dynamical Systems, P.** *Hamiltonian Dynamical Systems.* Ed: Kenneth R. Meyer, Donald G. Saari. Contemp. Math., V. 81. AMS, 1988, xiv + 270 pp, $28 (P). [ISBN: 0-8218-5086-5] Proceedings of a conference held June 1987 at the University of Colorado, Boulder. The sixteen papers presented here range from general expository to specific new results, and include topics as diverse as the $N$-body problem, KAM theory, twist and annulus maps, homoclinic and heteroclinic phenomena. CE

**Probability, T(12-13), S, C.** *Geometric Probability.* IBM PC. North Carolina School of Science and Mathematics. NCTM, 1988, v + 40 pp, $9 (P). [ISBN: 0-87353-259-7] The innovative opening section of a new senior course at the North Carolina School of Science and Mathematics: several extended problems in geometric probability with related exercises and associated simulation software on an IBM PC disk. An excellent introduction to modelling that both builds on traditional school topics and introduces ideas common to college mathematics. LAS

**Stochastic Processes, T(18: 1), P.** *Estimation, Control, and the Discrete Kalman Filter.* Donald E. Catlin. Appl. Math. Sci., V. 71. Springer-Verlag, 1989, xiii + 274 pp, $49.80. [ISBN: 0-387-96777-X] Wide-ranging book covering measure-theoretic probability, entropy, minimum variance estimation, adjoints and pseudoinverses in Hilbert spaces, and the Kalman filter. For the intellectually curious; not narrowly focused on how to use a Kalman filter. TH

**Stochastic Processes.** *General Theory of Markov Processes.* Michael Sharpe. Pure & Appl. Math., V. 133. Academic Pr, 1988, xi + 419 pp, $49.50. [ISBN: 0-12-639060-6] Reference to the theory of right processes. Extends work by Dellacherie, Meyer, and Getoor. Not for the inexpert. TH

**Statistics, T(17: 1, 2), S, P.** *Nonlinear Regression Analysis and Its Applications.* Douglas M. Bates, Donald G. Watts. Prob. & Math. Stat. Wiley, 1988, xiv + 365 pp, $39.95. [ISBN: 0-471-81643-4] Clear presentation of theory and practice of nonlinear regression. Extensive displays of geometrical constructs. Numerous examples using real data from physical, chemical, and biological experiments. Pseudocode for computing algorithms in three high-level languages: S, GAUSS, and SAS/IML. MS

**Statistics, P*.** *The Inverse Gaussian Distribution: Theory, Methodology, and Applications.* Raj S. Chhikara, J. Leroy Folks. Statistics, V. 95. Mar-

cel Dekker, 1989, viii + 213 pp, $79.75. [ISBN: 0-8247-7997-5] Summarizes known properties of the inverse Gaussian distribution, which arises as the distribution of the first passage time of a Brownian motion, and its uses in theoretical and applied statistics. Includes a wide variety of applications, particularly in the areas of life-testing and reliability. Extensive bibliography. Note price! RSK

**Statistics, T(17), S, P.** *Applied Time Series Analysis for Business and Economic Forecasting.* Sufi M. Nazem. Stat.: Textbooks & Mono., V. 93. Marcel Dekker, 1988, x + 431 pp, $89.75. [ISBN: 0-8247-7913-4] Specifically for those with modest background in mathematics and statistics. Covers philosophy and characteristics of time series. Key statistical measures for identifying and validating time series models, the analysis of simple models, and real-world applications using ARIMA. Many excellent examples. Can be used for self-teaching too. MS

**Statistics, T(16-17: 1), S.** *Lecture Notes in Statistics-46: Nonparametric Regression Analysis of Longitudinal Data.* Hans-Georg Müller. Springer-Verlag, 1988, vi + 199 pp, $23.40 (P). [ISBN: 0-387-96844-X] Reviews some work done for longitudinal data using basic mathematical tools that have been applied. Applications emphasized; Fortran programs provided. Good references included. MS

**Statistics, T?(18: 1), S*.** *Prescriptions for Working Statisticians.* Albert Madansky. Texts in Stat. Springer-Verlag, 1988, xix + 295 pp, $39.80. [ISBN: 0-387-96627-7] Excellent for applied statisticians. Techniques for inferences from specialized data, mixed categorical and measured data, and cross-classified data. Includes IDA and MINITAB examples; also Fortran programs. Background in matrix algebra and regression are helpful. Appropriate tables given throughout the text. MS

**Statistics, S(17).** *Asymptotics for Generalized Chi-Square Goodness-of-Fit Tests.* F.C. Drost. CWI Tract, V. 48. Mathematisch Centrum, 1988, iv + 104 pp, Dfl. 16.50 (P). [ISBN: 90-6196-348-6] Monograph considering behavior of various types of chi-square goodness-of-fit test statistics. Investigates influence of number of classes in the presence of a location-scale nuisance parameter; choice of location-scale estimator, and power approximations for Cressie-Read class of $\chi^2$ statistics when no nuisance parameters are present. MS

**Statistics, P*.** *Dynamic Graphics for Statistics.* Ed: William S. Cleveland, Marylyn E. McGill. Wadsworth, 1988, xii + 424 pp, $44.95. [ISBN:0-534-09144-X] Dynamic graphical methods involve direct manipulation of graphical elements on a computer graphics screen, with virtually instantaneous change of the elements. This is a collection of sixteen papers on the analysis of multidimensional data using these methods. Covering roughly a twenty-year time span, they also show how these ideas have evolved. RSK

**Programming, T(14: 1), L.** *Structured Programming in Assembly Language for the IBM PC.* William C. Runnion. Ser. in Comput. Sci. PWS-Kent, 1988, xvii + 697 pp. [ISBN: 0-534-91480-2] A text for course CS3 introducing computer organization via assembly language for the 8088 and 8086 microprocessors. (Special sections deal with differences applicable to the 80286 chips used in the AT machines.) A disk inside the back jacket contains source code and executable modules for example programs in the text along with object code for I/O and BCD subprocedures. LAS

**Computer Science, P, L**.** *Annual Review of Computer Science, Volume 3, 1988.* Ed: Joseph F. Traub, *et al.* Annual Reviews, 1988, 423 pp, $45. [ISBN: 0-8243-3203-2] Thirteen survey papers on a variety of topics from database security to artificial intelligence, from geometric computing to computational complexity. Cumulative author and title indices for Volumes 1-3. Excellent, high-quality papers make this series a superb resource for student and professional alike. LAS

**Applications, P, L.** *Mathematical Modelling in Science and Technology.* Ed: Ervin Y. Rodin, Xavier J.R. Avula. Math. & Computer Modelling, V. 11. Pergamon Pr, 1988, xvii + 1216 pp, (P). [ISBN: 0-08-036380-6] Proceedings of the Sixth International Conference on Mathematical Modelling in St. Louis in August, 1987. Over 250 papers on every conceivable type of mathematical modeling from chemical and biological systems to ocean acoustics and electro-optics, from cancer and neural models to cellular automata and strategic defense analysis. A stunning portrait of the diversity of mathematical modeling. LAS

**Applications (Engineering), T(18: 1), P.** *First-Order Partial Differential Equations, Volume II: Theory and Application of Hyperbolic Systems of Quasilinear Equations.* Hyun-Ku Rhee, Rutherford Aris, Neal R. Amundson. Intern. Ser. in Physic. & Chem. Engin. Sci. Prentice-Hall, 1989, xii + 548 pp. [ISBN: 0-13-319237-7] First order systems of partial differential equations arising primarily in chemical engineering are considered. Roughly, one-fourth of this book is general theory. The rest is concerned with such applications as $n$-solute chromatography, wave interactions, fixed and moving-bed adsorption. Includes exercises. SP

**Applications (Fluid Dynamics), P, L.** *Dynamics of Internal Layers and Diffusive Interfaces.* Paul C. Fife. CBMS-NSF Reg. Conf. Ser. in Appl. Math., V. 53. SIAM, 1988, vi + 93 pp, (P). [ISBN: 0-89871-225-4] Lectures from a May 1987 CBMS Regional Research Conference at Little Cottonwood Canyon, Utah on dynamics of internal layers, flame theory, electrophoresis, and waves in self-oscillatory media. Includes an extensive bibliography on a very timely topic. LAS

**Applications (Fluid Dynamics), P.** *Lecture Notes in Mathematics-1360: Vortex Methods.* Ed: C. Anderson, C. Greengard. Springer-Verlag, 1988, 141 pp, $13.10 (P). [ISBN: 0-387-50526-1] Proceedings of the UCLA Workshop on Vortex Methods held during May 1987. The ten research articles collected

here include numerical and theoretical investigations of incompressible fluid vorticity for two- and three-dimensional flow, convergence of vortex methods, and computationally-rapid methods. CE

**Applications (Fluid Dynamics), P.** *Mathematical Modelling of Ocean Circulation.* G.I. Marchuk, A.S. Sarkisyan. Springer-Verlag, 1988, xv + 292 pp, $120. [ISBN: 0-387-18925-4] Algorithms are presented and results discussed for modelling ocean dynamics. Influence of temperature, salinity, bottom topography, and shape of coast line are taken into account. SP

**Applications (Physics), S(18), P.** *The Theory of Pseudo-rigid Bodies.* Harley Cohen, Robert G. Muncaster. Tracts in Natural Philo., V. 33. Springer-Verlag, 1988, x + 183 pp, $69.80. [ISBN: 0-387-96635-8] Containing much previously unpublished material, this monograph represents a new approach to the theory of the deformation of bodies. The lucid writing contains many references to allied concerns, and is laced with fine mathematics. MU

**Applications (Physics), P.** *Nonlinear Evolution Equations.* J.K. Engelbrecht, V.E. Fridman, E.N. Pelinovski. Pitman Res. Notes in Math. Ser., V. 180. Longman Scientific & Technical (US Distr: Wiley), 1988, 122 pp, $44.95 (P). [ISBN: 0-582-01314-3] Three methods are developed to model nonlinear wave processes in weakly dispersive media: the iterative, asymptotic, and spectral methods. A comparative analysis of these methods is presented. Translated from the original Russian. SP

**Applications (Physics), S(18), P.** *Lecture Notes in Physics-313: Group Theoretical Methods in Physics.* Ed: H.-D. Doebner, J.-D. Hennig, T.D. Palev. Springer-Verlag, 1988, xi + 599 pp, $61.10. [ISBN: 0-387-50245-9] Proceedings of the XVI International Colloquium held at Varna, Bulgaria, June 1987. MU

**Applications (Physics), S(18), P.** *Special Relativity and Quantum Theory: A Collection of Papers on the Poincaré Group.* Ed: M.E. Noz, Y.S. Kim. Fundamental Theories of Physics. Kluwer Academic, 1988, xiii + 504 pp, $117. [ISBN: 90-277-2799-6] Based on E.P. Wigner's highly original 1939 paper "On Unitary Representations of the Inhomogeneous Lorentz Group," this volume is an extension of the textbook *Theory and Applications of the Poincaré Group,* and consists of Wigner's paper plus "the major papers on the Lorentz group published since 1939." MU

**Applications (Physics), S(18), P.** *Wave Phenomena: Theoretical, Computational, and Practical Aspects.* Ed: Lui Lam, Hedley C. Morris. Springer-Verlag, 1988, xii + 275 pp, $49.95. [ISBN: 0-387-96921-7] Proceedings of the First Woodward Conference held at San Jose State University, June 1988. Topics of a theoretical nature include: pseudodifferential operator techniques, inverse problems, and the mathematical foundations of wave propagation in random media. MU

**Applications (Physics), P*.** *Spinors and Space-Time, Volume 2: Spinor and Twistor Methods in Space-Time Geometry.* Roger Penrose, Wolfgang Rindler. Cambridge U Pr, 1986, ix + 501 pp, $34.50 (P). [ISBN: 0-521-34786-6] Introduction to the theory of twistors, mainly the way twistor theory interrelates with 2-component spinor theory. Topics covered include massless fields, the geometry of light rays, and the conformal structure of infinity. Twistorial applications abound and are presented in great detail. The approach here is to show how problems with the standard physical framework may be solved using new twistorial techniques. Includes a summary of *Volume 1* (TR, August-September 1988; 1986 hardcover edition, TR, December 1986). MR

**Applications (Physics), P.** *Differential Geometrical Methods in Theoretical Physics.* Ed: K. Bleuler, M. Werner. NATO ASI Ser. C, V. 250. Kluwer Academic, 1988, xvii + 471 pp, $108. [ISBN: 90-277-2820-8] Proceedings of the NATO Advanced Research Workshop and the 16th International Conference on Differential Geometric Methods in Theoretical Physics held in Como, Italy during August 1987. The twenty-five papers presented here are mostly research articles in quantum field theory, grouped broadly under the headings of string theory, integrable systems, symplectic structures, general relativity, and supersymmetry. CE

**Applications (Physics), T(15-17: 1), S, L.** *General Relativity: A Guide to its Consequences for Gravity and Cosmology.* J.L. Martin. Halsted Pr, 1988, xii + 176 pp, $49.95. [ISBN: 0-470-21183-0] The text presupposes only a modest background in mathematics, i.e., calculus and elementary differential equations. However, because of its unusual emphasis on the observational aspects of general relativity, little space is devoted to the development of the theory. It should prove very enjoyable to those already familiar with general relativity. MU

**Applications (Physics), S(18), P.** *Mark Kac Seminar on Probability and Physics: Syllabus 1985-1987.* Ed: F. den Hollander, H. Maassen. CWI Syllabus, V. 17. Mathematisch Centrum, 1988, 162 pp, Dfl. 25.30 (P). [ISBN: 90-6196-350-8] Lecture notes from a monthly seminar in Amsterdam where probabilists and statistical physicists discuss papers on topics of common interest. LAS

### Reviewers

SB: Steve Benson, St. Olaf; JNC: Judith N. Cederberg, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; CE: Christopher Ennis, Carleton; SG: Steven Galovich, Carleton; GG: George Gilbert, St. Olaf; RH: Rhonda Hatcher, St. Olaf; TH: Timothy Hesterberg, St. Olaf; JJ: Jason Jones, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; GN: Gail Nelson, Carleton; AO: Arnold Ostebee, St. Olaf; SP: Samuel Patterson, Carleton; MR: Matthew Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MS: Myriam Steinback, Macalester; MU: Milton Ulmer, Carleton; PZ: Paul Zorn, St. Olaf.

# GEOMETRY & TOPOLOGY

## FIRST CONCEPTS OF TOPOLOGY: The Geometry of Mappings of Segments, Curves, Circles, and Disks

W.G. Chinn and N.E. Steenrod

*This clear and winning little book, for readers willing to come to genuine grips with the idea of a mathematical proof, presents topology . . . as mathematicians see it. The parlor tricks are gracefully alluded to here and there, but they are distinctly for after hours. The center of interest is the stuff itself: powerful notions of set theory are . . . exploited to define open sets and their coverings, and from them [to prove] the key theorems. .. One cannot any longer doubt that a single stroke of a knife exists that divides any irregular ham sandwich so that ham and both bread slices can be shared with perfect fairness by two consumers.*

Philip Morrison in *Scientific American*

Contains over one hundred and fifty problems and solutions.

160 pp., 1966, ISBN 0-88385-618-2

List: $10.00 MAA Member: $8.00

Catalog Number NML-18

## FROM PYTHAGORAS TO EINSTEIN

K.O. Friedrichs

Starting with area and dissection proofs of the Pythagorean theorem, Friedrichs proceeds gently, step by step, to cover vectors, coordinates, elastic and inelastic impacts, and relativistic space-time–ending with a derivation of a contemporary formula rivaling the Pythagorean theorem in fame, $E = mc^2$.

88 pp., 1965, ISBN 0-88385-616-6

List: $9.50 MAA Member: $7.60

Catalog Number NML-16

## GEOMETRY REVISITED

H.S.M. Coxeter and S.L. Greitzer

After seeing a proof that a segment has a unique midpoint, students may wonder whether there are any interesting theorems in geometry. If either they or their teachers get a hold of this book, they will learn that there are many beautiful and nontrivial theorems in geometry. Among those found here are the theorems of Ceva, Menelaus, Pappus, Desargues, Pascal, and Brianchon. A nice proof is given of Morley's remarkable theorem on angle trisectors. The transformational point of view is emphasized: reflections, rotations, translations, similarities, inversions, and affine and projective transformations. Many fascinating properties of circles, triangles, quadrilaterals, and conics are developed.

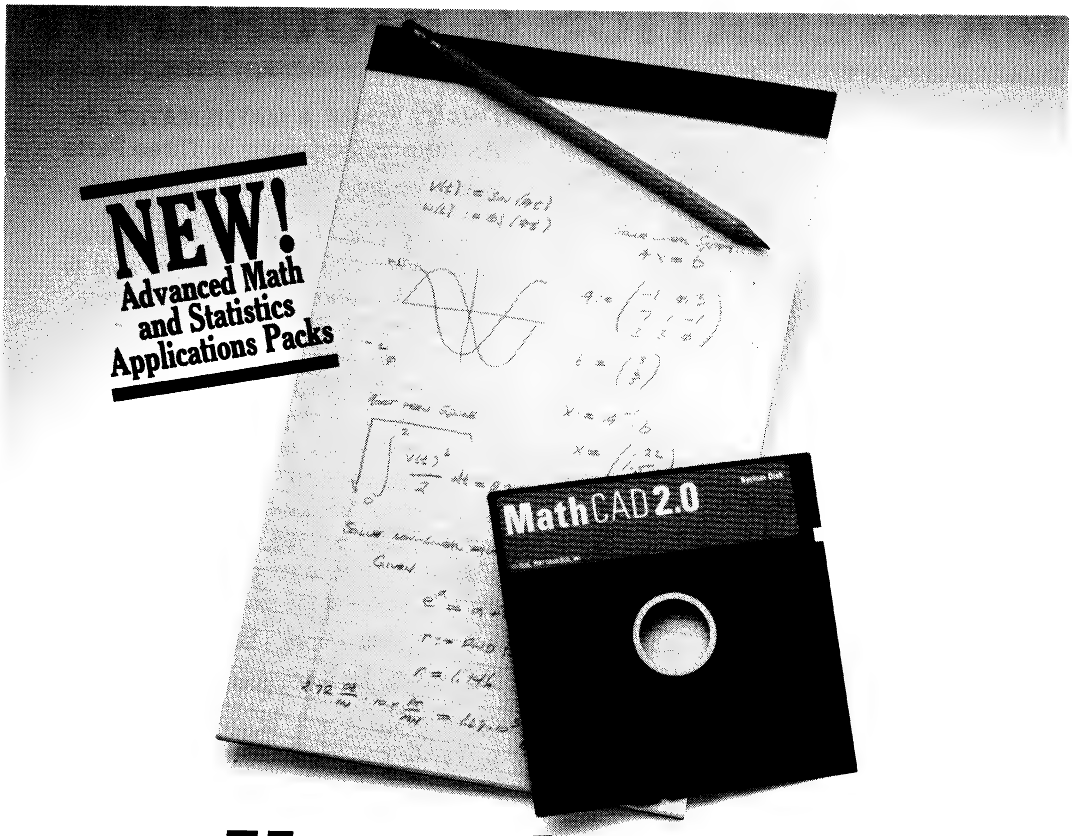
193 pp., 1967, ISBN 0-88385-619-0

List: $11.75 MAA Member: $9.50

Catalog Number NML-19

*The very lucid presentation takes the reader from the elementary problems of plane Euclidean geometry to the fundamental concepts of non-Euclidean geometry . . . The book is rich in remarkable facts and thereby is very effective in promoting the significance and the value of geometry in mathematical teaching, a promotion which is very necessary . . . The always original developments use very simple tools . . . and soon proceed to higher configurations.*

K. Strubecker in *Mathematical Reviews*



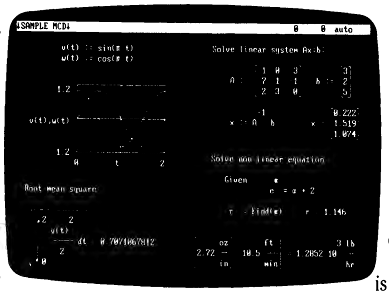

# Your pad or ours?

If you perform calculations, the answer is obvious.

MathCAD 2.0.

It's everything you appreciate about working on a scratchpad – simple, free-form math – and more. More speed. More accuracy. More flexibility.

Just define your variables and enter your formulas anywhere on the screen. MathCAD formats your equations as they're typed. Instantly calculates the results. And displays them exactly as you're used to seeing them – in real math notation, as numbers, tables or graphs.




MathCAD is more than an equation solver. Like a scratchpad, it allows you to add text anywhere to support your work, and see and record every step. You can try an unlimited number of what-ifs. And print your entire calculation as an integrated document that anyone can understand.

Plus, MathCAD is loaded with powerful built-in features. In addition to the usual trigonometric and exponential functions, it includes built-in statistical functions, cubic splines, Fourier transforms, and more. It also handles complex numbers and unit conversions in a completely transparent way.

Yet, MathCAD is so easy to learn, you'll be using its full power an hour after you begin.

What more could you ask for? How about two new applications packs to increase your productivity?

The **Advanced Math Applications Pack** includes 16 applications like eigenvalues and eigenvectors of a symmetric matrix, solutions of differential equations, and polynomial least-squares fit.

The **Statistics Applications Pack** lets you perform 20 standard statistical routines such as multiple linear regression, combinations and permutations, finding the median, simulating a queue, frequency distributions, and much more.

MathCAD lets you perform calculations in a way that's faster, more natural, and less error-prone than the way you're doing them now – whether you use a calculator, a spreadsheet, or programs you write yourself. So come on over to MathCAD and join 45,000 enthusiastic users.

For more information, contact your dealer or call **1-800-MATHCAD** (In MA: 617-577-1017).

Requires IBM PC® or compatible, 512KB RAM, graphics card.

IBM PC® International Business Machines Corporation
MathCAD® MathSoft, Inc

**MathCAD®**

**MathSoft, Inc., One Kendall Sq., Cambridge, MA 02139**

# Winner of the First MAA Book Prize . . .


Carus Mathematical Monograph, #19

## Field Theory and its Classical Problems,

by Charles R. Hadlock

323 + xvi pp. Hardbound

List: $28.00 MAA Member: $22.00

**Field Theory and its Classical Problems** is one of those rare monographs that will capture and hold a mathematically prepared reader (one reasonably familiar with basic calculus and linear algebra) until the last page. Mathematicians will be intrigued by the development which places Galois Theory in an historical perspective and allows it to unfold from the emergence of the classical construction problems to the discovery of their solutions. The book presents some of history's greatest and most elegant mathematics.

Read what some reviewers say about this monograph:

> *"Hadlock has produced a pedagogic masterpiece . . . His idea of beginning with the three Greek problems and letting them fire the reader's spark plugs is brilliant . . . (his) ability to inject vitality and enthusiasm into mathematical text is remarkable.* (George Piranian, University of Michigan)

> *"In the preface, the author claims to have written the book for himself . . . Fortunately for us, he chose to share his work with the mathematical community. I suggested the book as collateral reading in a one-semester course in Galois Theory and the students found it very readable and most helpful in establishing a motivation and direction . . . This is a delightful book for both student and teacher."* (John D. Leadley, ***Mathematical Reviews***)

**Table of Contents**

**Chapter 1**—The Three Greek Problems
Constructible Lengths; Doubling the Cube; Trisecting the Angle; Squaring the Circle; Polynomials and Their Roots; Symmetric Functions; the Transcendence of $\pi$.

**Chapter 2**—Field Extensions
Arithmetic of Polynomials; Simple, Multiple, and Finite Extensions; Geometric Constructions Revisited; Roots of Complex Numbers; Constructibility of Regular Polygons I; Congruences; Constructibility of Regular Polygons II.

**Chapter 3**—Solution by Radicals
Statement of the Problem; Automorphisms and Groups; The Group of an Extension; Two Fundamental Theorems, Galois' Theorem; Abel's Theorem; Some Solvable Equations.

**Chapter 4**—Polynomials With Symmetric Groups
Background Information; Hilbert's Irreducibility Theorem; Existence of Polynomials over Q with Group $S_n$.

Order from:
**The Mathematical Association of America**
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

# THREE IMPORTANT BOOKS ON MATHEMATICS EDUCATION

## RESPONSES TO THE CHALLENGE: KEYS TO IMPROVED INSTRUCTION BY TEACHING ASSISTANTS AND PART-TIME INSTRUCTORS

The Committee on Teaching Assistants
and Part-Time Instructors,
Bettye Anne Case, Chair

The committee that prepared this volume has been gathering information on policies, practices, successes, failures, and goals connected with the use of teaching assistants and part-time instructors. In this volume the committee presents and analyzes data showing who these teachers are, the extent and nature of their teaching duties, and the efforts made to assimilate them into the faculties. This volume will help you to see how your department compares nationally, to decide what steps you and your school should take, and to understand what additional resources might be needed.

280 pp., 1988, ISBN-0-88385-061-3
List: $15.00
Catalog Number NTE-11

## GUIDELINES FOR THE CONTINUING MATHEMATICAL EDUCATION OF TEACHERS

Committee on the Mathematical Education of Teachers

These guidelines will be very useful to school teachers and supervisors, to college administrators who plan continuing education programs, to the college teachers who design and teach courses for teachers, and to school administrators who must think about requirements for continuing education of teachers. The guidelines are rich in specifics on course content, giving clear objectives for all courses. Teachers who want to dig out material for themselves or in order to enrich their classes will find the more than 500 references provided here under various topics an invaluable aid.

90 pp., 1988, ISBN-0-88385-060-5
List: $8.00
Catalog Number NTE-10

## THE USE OF CALCULATORS IN THE STANDARDIZED TESTING OF MATHEMATICS

John W. Kenelly, Editor

The calculator is a universal tool for all those involved in quantitative work from science and engineering to business. Routine use of calculators is part of the training and testing of students headed for these fields. But this is not yet the case in mathematics. This symposium, jointly sponsored by the MAA and The College Board, sets out clearly the theoretical and practical issues that must be addressed as calculators are brought more fully into the mathematics curriculum. This is a practical group concerned with specific tests and test items. General theoretical considerations are set off by the specifics of individual test items and students' success rates on them. The Ohio Early College Mathematics Placement Test is reported on in detail by Joan R. Leitzel and Bert K. Waits. James W. Wilson and Jeremy Kilpatrick examine the theoretical issues in the development of calculator-based tests. John Harvey, now Chair of the MAA's Committee on Placement Examinations looks at the issues surrounding calculator use on placement examinations, as well as giving an overview of the symposium and a survey of developments through 1988.

vi + 50 pp., 1989, Copublished by the MAA and The College Board. LC No. 88-064-100

List: $6.50 Member: $6.50

*Order from:*

The Mathematical Association of America
1529 Eighteenth Street, N. W.
Washington, D. C. 20036
(202) 387-5200

# MAA STUDIES IN MATHEMATICS

## Studies in Combinatorics

Gian-Carlo Rota, editor
Volume #17, MAA Studies in Mathematics
272 pp., 1978, Hardbound, ISBN-0-88385-177-2
List: $27.50 MAA Member: $21.50

*"This is a gem of a book. In layout and size it brings back memories of the classic* **Combinatorial Mathematics** *by H.J. Ryser. The index is very good and the editor is to be warmly congratulated on bringing out a book of enormous value and uniformly high standard. It is the sort of book which could be used with great profit as the basis of a weekly working seminar on combinatorics. I recommend it unreservedly."*

D.J.A. Welsh in ***Bulletin of the Institute of Mathematics and its Applications***

This excellent book in the **MAA Studies in Mathematics** series contains seven papers by prominent researchers in combinatorics. Although written by different authors, the papers have the unity of chapters in a monograph. Together they present a fairly comprehensive exposition of some recent developments in combinatorics. Much of the material has application outside mathematics to computer science, operations research and network and circuit theory. Begun at the time of Euler, combinatorics has become one of the most active fields in mathematics. A significant feature of **Studies in Combinatorics** is its accessibility to any mathematician qualified to teach at the college level, as well as to a great many students. This volume is also accessible to a wide scientific public outside mathematics.

**Table of Contents**


Combinatorial Matrix Theory . . . . . . . . H.J. Ryser
Proof Techniques in the Theory of Finite Sets . . . Curtis Greene and Daniel J. Kleitman
Ramsey Theory . . . . . . . . R.L. Graham and B.L. Rothschild
Generating Functions . . . . . . . . Richard P. Stanley
Nonconstructive Methods in Discrete Mathematics . . . . . . . . Joel Spencer
Matroids and Combinatorial Geometries . . . . . . Tom Brylawski and Douglas G. Kelly
Combinatorial Constructions . . . . . . . . Marshall Hall, Jr


Order From:
**The Mathematical Association of America**
1529 Eighteenth Street, N.W.
Washington, D.C. 20036

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N W
Washington, DC 20036

THE AMERICAN
# MATHEMATICAL MONTHLY

Volume 96, Number 6 June July 1989

## Contents

(ISSN 0002-9890)

ARTICLES

Disks, Balls, and Walls: Analysis of a Combinatorial Game . . . . . . . . . . . . . . RICHARD ANDERSON, LÁSZLÓ LOVÁSZ, PETER SHOR, JOEL SPENCER, EVA TARDOS, AND SHMUEL WINOGRAD 481

Mountain Climbing, Ladder Moving, and the Ring-Width of a Polygon . . . . . . . . . . JACOB E. GOODMAN, JÁNOS PACH, AND CHEE K. YAP 494

LETTERS TO THE EDITOR . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 510

NOTES

Monotone Multiplicative Functions . . . . . . . . . . . . . . . . . . . . . . JOEL M. COHEN 512

Powers of a Prime Dividing Binomial Coefficients . . . . . . . . . . . . . . . W. J. WONG 513

The Polar Decomposition and a Matrix Inequality . . . . . . . . . . . DERMING WANG 517

THE TEACHING OF MATHEMATICS

A Natural Interpretation of an Artificial Function . . . . . . HANSKLAUS RUMMLER 520

Factor Rings of Integers . . . . . . . . . . . . . . . . . . . . . . . . . . . . STEVE JOHNSON 521

A Simple Estimate of the Error in Linear Approximation . . ROBERT M. GETHNER 522

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 524

Advanced Problems and Solutions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 533

REVIEWS

Mathematics with Applications, Fourth Edition. By Margaret L. Lial and Charles D. Miller . . . . . . . . . . . . CLINTON J. OXENRIDER 537

Sphere Packings, Lattices and Groups. By J. H. Conway and N. J. A. Sloane . . . . . . . . . . . . . . . . . . H. S. M. COXETER 538

Geometric Inequalities. By Yu. D. Burago and V. A. Zalgaller . . . . . . . . . . . . . . . . . . DON CHAKERIAN 544

TELEGRAPHIC REVIEWS . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 547

# NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to ANITA E. SOLOW, Department of Mathematics, Grinnell College, Grinnell, IA 50112; in geometry or topology to DENNIS DETURCK, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Three copies of proposed problems or solutions should be sent to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, University of Calgary, Alberta, Canada T2N 1N4




Reprint Permission: A. B. WILLCOX, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: MS. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $26 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $29. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $90 per year.





Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Assocation of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

# Disks, Balls, and Walls: Analysis of a Combinatorial Game

RICHARD ANDERSON, *University of Washington, Seattle, WA*
LÁSZLÓ LOVÁSZ, *Eötvös Loránd University and Princeton University*
PETER SHOR, *AT & T Bell Labs, Murray Hill, NJ*
JOEL SPENCER, *Courant Institute, NYU, New York City*
EVA TARDOS, *Massachusetts Institute of Technology, Cambridge, MA*
SHMUEL WINOGRAD, *IBM, T. J. Watson Research Center, Yorktown Heights, NY*

Richard Anderson received his Ph.D. from Stanford University in 1986. He is now an Assistant Professor of Computer Science at the University of Washington.

László Lovász received his Dr.Rher.Nat. degree from the Eötvös Loránd University in Budapest in 1971. Currently he is Professor of Computer Science at the Eötvös Loránd University and at Princeton University.

Peter Shor received his Ph.D. from MIT in 1985. He is currently working at AT & T Bell Labs in Murray Hill. He has most recently been doing research in computational geometry.

Joel Spencer is Professor of Mathematics and Computer Science at the Courant Institute. His research interests are in combinatorial analysis, particularly Ramsey theory and the probabilistic method.

Eva Tardos received her degree from Eötvös University in Budapest. She is currently an Assistant Professor of Applied Mathematics at Massachusetts Institute of Technology.

Shmuel Winograd received his Ph.D in mathematics from NYU in 1968. He is currently the Director of the Mathematical Sciences Department at IBM, T. J. Watson Research Center.

**1. Introduction.** In this article we study a very simple "game" that can be played with several piles of disks arranged in a line. At each unit of time, each pile is divided in half, with half the disks being added to the pile on the left and the other half added to the pile on the right. If there is an odd number of disks in a pile, one

*This work was done at the Mathematical Sciences Research Institute, Berkeley, California.

disk is left stationary and the remainder is divided evenly. It is easy to simulate this game on paper for small numbers of disks. This provides an interesting diversion during long lectures or faculty meetings. The process is also easy to program on a personal computer. Some of the phenomena we describe are best seen "live" on a computer terminal.

We are interested in investigating what happens when we start out with a single pile of disks. It is easy to work out small examples by hand. For example, with 5 disks we get

$$\begin{array}{ccccc}
 & & 5 & & \\
 & 2 & 1 & 2 & \\
1 & 0 & 3 & 0 & 1 \\
1 & 1 & 1 & 1 & 1
\end{array}$$

After three moves we get 5 consecutive piles of one disk each, so no further moves may be made. A bigger example yields a similar result:

$$\begin{array}{ccccccccc}
 & & & & 9 & & & & \\
 & & & 4 & 1 & 4 & & & \\
 & & 2 & 0 & 5 & 0 & 2 & & \\
 & 1 & 0 & 3 & 1 & 3 & 0 & 1 & \\
 & 1 & 1 & 1 & 3 & 1 & 1 & 1 & \\
 & 1 & 1 & 2 & 1 & 2 & 1 & 1 & \\
 & 1 & 2 & 0 & 3 & 0 & 2 & 1 & \\
 & 2 & 0 & 2 & 1 & 2 & 0 & 2 & \\
1 & 0 & 2 & 0 & 3 & 0 & 2 & 0 & 1 \\
1 & 1 & 0 & 2 & 1 & 2 & 0 & 1 & 1 \\
1 & 1 & 1 & 0 & 3 & 0 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{array}$$

On the basis of examples that can be worked out by hand, it is natural to conjecture that if the starting position is a single pile with $2n + 1$ disks, then the final configuration is $2n + 1$ consecutive piles with one disk each. Our first result is that, in fact, that is the case. The next problem is to determine how many moves it takes to reach the final configuration. We are able to bound the number of moves taken when starting with $2n + 1$ disks to between $\alpha n^2$ and $\beta n^2$. In working out larger examples, many patterns start to appear. It turns out that there is a relationship between the patterns and certain differential equations which lead to intriguing interpretations of the problem.

We can define the "game" mathematically by letting $a_{i,t}$ be the number of disks in pile $i$ at time $t$. The initial conditions are given by the value of $a_{i,0}$. We are primarily interested in the case where $a_{0,0} = m$ and $a_{i,0} = 0$ for $i \neq 0$. The value of $a_{i,t}$ for $t \geqslant 1$ is given by

$$a_{i,t} = \left\lfloor \frac{a_{i-1,t-1}}{2} \right\rfloor + \left\lfloor \frac{a_{i+1,t-1}}{2} \right\rfloor + \left( a_{i,t-1} \bmod 2 \right).$$

As long as some pile has at least 2 disks, moves may be made. The final configuration is the one where all piles have at most one disk.

To study the disk moving problem, we look at a seemingly more general game. In this new game we also have piles of disks arranged in a line. However, a move is to pick some pile with at least 2 disks and move one disk to the left and another to the

right. There will often be more than one move available at a time. There is a sequence of moves in the new game that has the same effect as a single move in the original game, so the original game can be simulated by the new game. However, it is conceivable that the final position for the new game might depend on the order in which the moves are made. We shall show that the order, which does not matter, will allow us to use the new game to study the original one.

**2. Final position.** We begin by proving a few simple properties of the sequence of moves and the final configuration. In particular, we must show that the game does reach a position where no more moves can be made.

For the remainder of the section we assume that the input configuration is $2n + 1$ disks at the origin.

LEMMA 1. *It is not possible to have two consecutive empty piles which have nonempty piles on either side.*

*Proof.* Suppose piles $x$ and $x + 1$ are empty and there are disks both to their left and right. Both $x$ and $x + 1$ had to have been occupied at some time. Suppose $x + 1$ was the last occupied of the two. That means disks were moved off of it, and one of them would have gone to $x$. Hence, $x$ was occupied after $x + 1$ was, contradiction. ∎

LEMMA 2. *The only piles that could receive disks are those numbered between* $-4n$ *and* $4n$.

*Proof.* Since the maximum gap between nonempty piles is one, the $2n + 1$ disks can be spread out over a range of at most $4n + 1$. After the first move is made, there will always be disks on both side of pile zero, so the range must be inside of $[-4n, 4n]$. ∎

Now that we have shown the range is bounded, we also show that the number of moves is bounded.

LEMMA 3. *The number of moves is bounded by* $32n^3 + 16n^2$.

*Proof.* Let $a_i$ be the number of disks in the $a_i$th pile, for each $i$, and consider $\Sigma_i a_i i^2$, the sum of the squares of the distances from the origin. The maximum value of this sum is if all disks were at $-4n$ and $4n$, which would give a sum of $32n^3 + 16n^2$. Every move increases the sum by exactly 2. A move from the pile $i$ adds elements at $i - 1$ and $i + 1$ and removes two elements from $i$. The change to the sum is $(i + 1)^2 + (i - 1)^2 - 2i^2 = 2$. Thus, there could not be more than $16n^3 + 8n^2$ moves or else the sum would exceed its maximum. ∎

THEOREM 1. *The final position is independent of the order that moves are applied.*

*Proof.* The key to the proof is that if two moves can be applied at the same time, then they may be applied in either order. Suppose that positions $x$ and $y$ both have at least two disks. If $x$ and $y$ differ by more than one, then making moves in the order $xy$ or $yx$ clearly have the same result. The only case where there is a possibility for interaction between the moves is if they differ by one, so $x = y + 1$ or $x = y - 1$. However for either order, $xy$ or $yx$, the effect is to reduce the piles at

$x$ and $y$ by one and increase the two bounding piles each by one.

$$\begin{array}{ccccccccc}
 & & & & & +1 & -2 & +1 & \\
\underline{\quad} & \underline{\quad} & \underline{\quad} & \underline{\quad} & \xRightarrow{x} & \underline{\quad} & \underline{\quad} & \underline{\quad} & \underline{\quad} \\
 & x & y & & & & x & y & \\
 & & & & & & & & \\
 & & & & & +1 & -1 & -1 & +1 \\
 & & & & \xRightarrow{y} & \underline{\quad} & \underline{\quad} & \underline{\quad} & \underline{\quad} \\
 & & & & & & x & y &
\end{array}$$

$$\begin{array}{ccccccccc}
 & & & & & & +1 & -2 & +1 \\
\underline{\quad} & \underline{\quad} & \underline{\quad} & \underline{\quad} & \xRightarrow{y} & \underline{\quad} & \underline{\quad} & \underline{\quad} & \underline{\quad} \\
 & x & y & & & & x & y & \\
 & & & & & & & & \\
 & & & & & +1 & -1 & -1 & +1 \\
 & & & & \xRightarrow{x} & \underline{\quad} & \underline{\quad} & \underline{\quad} & \underline{\quad} \\
 & & & & & & x & y &
\end{array}$$

Now we show that the fact that moves commute implies that there is a unique final configuration. A configuration $B$ is a *descendant* of configuration $A$ if there is a sequence of moves that transforms configuration $A$ into configuration $B$. A configuration is *final* if no moves apply to it. Suppose that there exists at least two final configurations. Let $A$ be a configuration that has two final configurations among its descendants, but every proper descendant of $A$ has a single final configuration among its descendants. Suppose $A'$ and $A''$ are direct descendants of $A$ with different final descendants. It follows that $A'$ and $A''$ cannot have any common descendants. Let $A'$ be derived by making move $x$ from $A$ and $A''$ be derived by making move $y$. Consider the configuration $B$ formed by making moves $xy$ from $A$ (or equivalently moves $yx$ from $A$). Since $B$ is a descendant of both $A'$ and $A''$, we have a contradiction. ■

With the previous theorem in hand, we can establish the final position for a given input. If we start with $2n + 1$ disks at the origin, the final configuration is $2n + 1$ consecutive piles of size one occupying positions $-n$ through $n$. To prove this we can look at a particular sequence of moves that is easy to analyze.

THEOREM 2. *If we start with a pile of size $2n + 1$ at position zero, then the final position is piles of size one in positions $-n$ through $n$.*

*Proof.* We consider a sequence of moves which yield the following intermediate configurations:

$$\begin{array}{ccccccc}
 & & & 2n+1 & & & \\
 & & 1 & 2n-1 & 1 & & \\
 & 1 & 1 & 2n-3 & 1 & 1 & \\
1 & 1 & 1 & 2n-5 & 1 & 1 & 1 \\
 & & & \vdots & & &
\end{array}$$

Suppose we reach the position where we have $j$ consecutive 1's, 3 in position 0 and $j$ more consecutive 1's. At each step we perform all legal moves simultaneously.

The pattern is

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 1 | 1 | 1 | 3 | 1 | 1 | 1 | 1 | |
| | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 1 | |
| | 1 | 1 | 2 | 0 | 3 | 0 | 2 | 1 | 1 | |
| | 1 | 2 | 0 | 2 | 1 | 2 | 0 | 2 | 1 | |
| | 2 | 0 | 2 | 0 | 3 | 0 | 2 | 0 | 2 | |
| 1 | 0 | 2 | 0 | 2 | 1 | 2 | 0 | 2 | 0 | 1 |
| 1 | 1 | 0 | 2 | 0 | 3 | 0 | 2 | 0 | 1 | 1 |
| 1 | 1 | 1 | 0 | 2 | 1 | 2 | 0 | 1 | 1 | 1 |
| 1 | 1 | 1 | 1 | 0 | 3 | 0 | 1 | 1 | 1 | 1 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

In the configuration there is a region of consecutive ones and a region of alternating 2's and 0's. The 2's and 0's move to the end of the list causing it to increase in length and then retreat leaving a string of consecutive 1's.

We can now reduce a configuration of $j$ 1's, $2n + 1 - 2j$ at position zero followed by $j$ 1's to $j + 1$ 1's, $2n + 1 - 2(j + 1)$ at position zero, followed by $j + 1$ 1's, by considering all but 3 of the disks in pile zero fixed. By doing this repeatedly we can go from $2n + 1$ in position zero down to $2n + 1$ consecutive ones. ■

We have just considered the case where we start with an odd number of disks. Note that in this sequence of moves, one of the disks in pile zero never moves. Thus we can cover the even case by just ignoring the one disk that always is in position one. If we start with $2n$ disks, we end with $n$ consecutive disks in positions $-n$ through $-1$, position 0 empty, and $n$ more consecutive disks in positions 1 through $n$.

**3. Run time.** Now that we know what the final configuration is, we return to the original game where all moves are made simultaneously. We investigate how long it takes to reach the final configuration. We first give two simple arguments that establish a lower bound of $n^2/3$ and an upper bound of $n^2$ for the number of steps needed to reach the final configuration, starting with $2n + 1$ disks at position 0.

LEMMA 4. *If the initial configuration is $2n + 1$ disks at position zero, the number of steps until the final configuration is reached is at least $n^2/3$.*

*Proof.* As in Lemma 3, we consider the sum $\Sigma_i a_i i^2$. The value of the sum at the end is just $n(n + 1)(2n + 1)/3$. Each pair of elements that is moved increases the sum by two, so the sum can increase by at most $2n + 1$ in a single move. Hence the number of moves is at least $n(n + 1)/3$. ■

We now give an upper bound that shows that the final configuration is reached within $n^2$ steps. To prove the bound, we need a little more information about the number of disks in the piles. We first note that at an even time step, moves are made only from even numbered piles, and at an odd time step, moves are made only from odd numbered piles. We refer to the piles from which moves are made as *active*. A typical configuration midway through the games is something like this:

1 0 2 1 3 0 6 1 7 0 9 0 7 1 6 0 3 1 2 0 1

The number of disks per pile, decreases according to distance from the origin. The

formal condition that the number of disks per pile satisfies is

**Proposition** 1. *If the time is even, then the sequences* $a_0, a_2, a_4, \ldots$ *and* $a_0, a_{-2}, a_{-4}, \ldots$ *are nonincreasing. If the time is odd, then the sequences* $a_1, a_3, a_5, \ldots$ *and* $a_{-1}, a_{-3}, a_{-5}, \ldots$ *are nonincreasing.*

We defer the proof of this proposition. An immediate corollary to this is that the pile zero cannot be down to a single item at an even time step, before the final configuration is reached.

**Lemma** 5. *If the initial configuration is* $2n + 1$ *disks at position zero, the final configuration is reached within* $n^2 + n$ *moves.*

*Proof.* Here we consider the sum $\Sigma_i a_i|i|$. The final value of the sum is just $n^2 + n$. The only moves that increase the sum are those made from position 0. At every even time step, at least two disks are moved off pile zero, so the sum increases by at least two. Thus, the sum must exceed $n^2 + n$ by time $n^2 + n$. ∎

**Lemma** 6. *The number of disks in the active piles is decreasing with distance from the origin.*

*Proof.* We shall prove the lemma only for piles numbered greater than or equal to zero, since the other side of the origin follows by symmetry. The proof is by induction. We show that if the proposition is true at a given time, then it is still true at the next unit of time. The key to this proof is to have the proper induction hypothesis, which is actually stronger than the result that we seek for the lemma. What we show is that for $i \geqslant 0$, if $i$ and $i + 2$ are active at time $t$, then $a_{i,t} \geqslant a_{i+2,t}$ and if $a_{i,t} = a_{i+2,t}$ then $a_{i+1,t}$ agrees in parity with $a_{i,t}$ and $a_{i+2,t}$. Thus, 8, 0, 8 and 9, 1, 9 are acceptable sequences, but 8, 1, 8, and 9, 0, 9 are not. (Note that since $a_{0,t}$ is always odd, we do not require that $a_{0,t}$ agrees in parity with $a_{-1,t}$ and $a_{1,t}$.)

The result clearly holds at time 0 when the only nonempty pile is pile 0. Suppose the result holds at time $t - 1$; we shall show it holds at time $t$. We argue by contradiction, so suppose $i$ and $i + 2$ are active piles at time $t$, but $a_{i,t} < a_{i+2,t}$. Since $a_{i-1,t-1} \geqslant a_{i+3,t-1}$, the amount $i$ received from $i - 1$ is at least as great as the amount that $i + 2$ received from $i + 3$. The piles $i$ and $i + 2$ received an equal amount from $i + 1$. Thus, the only way for $a_{i,t} < a_{i+2,t}$ would be if $a_{i,t-1}$ was 0 and $a_{i+2,t-1}$ was 1. It must have been the case that either $a_{i-1,t-1} = a_{i+3,t-1}$ or $a_{i-1,t-1} = a_{i+3,t-1} + 1$ for $i$ to receive from $i - 1$ the same as $i + 2$ received from $i + 3$. In both cases, as long as $i \neq 0$, we have a violation of the parity condition at time $t - 1$. Since we have shown that $a_{i,t-1} = 0$, we have also ruled out the case when $i = 0$. Hence we can conclude that $a_{i,t} \geqslant a_{i+2,t}$.

Now we must show that the parity condition also holds at time $t$. Suppose that it does not hold. We consider the two cases where it could be violated.

1) $a_{i,t} = a_{i+2,t}$, $a_{i,t}$ is even and $a_{i+1,t} = 1$.

For this to occur, it must be that $a_{i+1,t-1} = a_{i,t} + 1$. But then $a_{i-1,t-1} = a_{i+1,t-1}$, or else $a_{i,t}$ would have been bigger. Thus, it must be the case that $a_{i,t} = 0$. This is not a possible value for $i = 0$, and is a violation of parity for $i > 0$.

2) $a_{i,t} = a_{i+2,t}$, $a_{i,t}$ is odd and $a_{i+1,t} = 0$. In this case, we can conclude that $a_{i+1,t-1} = a_{i,t} - 1$. However, this forces, $a_{i+3,t-1} = a_{i+1,t-1}$ and $a_{i+2,t-1} = 1$, a violation of parity.

Thus in either case, a violation of parity at time $t$ can be traced back to a violation of parity at time $t-1$. ■

We now give a better upper bound on the number of moves that the game takes. This upper bound is also based on the sum $\Sigma_i a_i |i|$. It relies on a more careful analysis of the sum's rate of growth to get a better bound.

**Theorem 3.** *If the initial configuration is* $2n+1$ *disks at position zero, the number of steps until the final configuration is reached is at most* $\left(\frac{\pi^2}{6}-1\right)n^2+O(n)$.

*Proof.* Let $S^1 = \Sigma_{0 \leqslant i \leqslant n} a_i i$. Initially $S^1 = 0$ and at the end, $S^1 = (1/2)(n^2+n)$. The change in $S^1$ at time $t$ is $\lfloor a_{0,t}/2 \rfloor$. Thus, the larger $a_0$ is, the faster $S^1$ grows. We get an upper bound on the total time by showing that the sum must be growing at a reasonable rate.

There is a smallest possible value (subject to the nonincreasing constraint for the piles) of $S^1$ such that $a_0 = 2k-1$. This is the value of $S^1$ if the disks are spread out "evenly" using as few piles as possible. So $a_0 = a_2 = \cdots = a_{2j} = 2k-1$, $a_1 = a_3 = \cdots = a_{2j+1} = 1$, and $a_{2j+2} = r$ for some $r \leqslant 2k-1$. We denote this value of $S^1$ as $\hat{S}$. The value of $j$ is $j = \left\lfloor \frac{n-k}{2k} \right\rfloor$.

$$\begin{aligned}
\hat{S} &= \sum_{i=0}^{j}(2i+1) + (2k-1)\sum_{i=1}^{j} 2j + (2j+2)r \\
&= 2kj^2 + (2k+2r+1)j + 2r + 1 \\
&\geqslant 2k\left(\frac{n-k}{2k}-1\right)^2 \\
&\geqslant \frac{n^2}{2k} - 3n.
\end{aligned}$$

Thus, if $S^1 < (n^2/2k) - 3n$, then $a_0 \geqslant 2k+1$. While $n^2/(2k+2) - 3n \leqslant S^1 < (n^2/2k) - 3n$, $S^1$ increases by at least $k$ every 2 steps, so the time to go between $n^2/(2k+2) - 3n$ to $(n^2/2k) - 3n$ is at most

$$n^2\left(\frac{1}{2k} - \frac{1}{2k+2}\right)\frac{2}{k} = \frac{n^2}{k^2(k+1)}.$$

We can get an upper bound by summing over $k$. We must add $7n$ since we have not accounted for the range $(n^2/2) - 3n$ to $(n^2/2) + (n/2)$. Thus our bound is

$$\begin{aligned}
n^2 \sum_{k \geqslant 1} \frac{1}{k^2(k+1)} + 7n &= n^2 \sum_{k \geqslant 1}\left(\frac{1}{k^2} - \frac{1}{k(k+1)}\right) + 7n \\
&= \left(\frac{\pi^2}{6} - 1\right)n^2 + 7n < .65n^2 + 7n.
\end{aligned}$$

■

## 4. A physical interpretation.

We now express a relationship between this game and a differential equation. We will be able to use this relationship to show that as $n$ goes to infinity, the number of moves is $cn^2 + o(n^2)$ for some fixed constant $c$. The

relationship can also be used to derive better bounds on this constant $c$ than we give in this paper. This differential equation is obtained by analyzing a pattern that appears after approximately $n$ steps of the game. By approximating the behavior of this pattern with a differential equation, we can get an approximation of the number of steps that gets better as $n$ increases.

The pattern can be described quite simply. Although it is difficult to define the exact point at which it begins, it cannot appear until the difference in the number of disks between adjacent active piles becomes at most 3, and it does appear shortly thereafter. Recall that Lemma 6 says that the number of disks in the active piles decreases as the piles get farther from the origin. When the number of disks at the origin becomes small, then there must be many piles which contain the same number of disks. The proof of Lemma 6 shows that if two adjacent active piles contain the same number of disks, then the (inactive) pile between them contains one or no disks depending on whether the two adjacent piles have an odd or even number of disks. Thus, there must be long regions having active piles that all contain the same value, so that a typical position will look like

$$\ldots, 2k+1, 1, 2k+1, 1, 2k+1, \ldots, 2k, 0, 2k, 0, 2k, \ldots,$$
$$2k-1, 1, 2k-1, 1, 2k-1, \ldots$$

It is easy to check that a region where piles alternate between $2k+1$ and 1 (or $2k$ and 0) remains unchanged; thus all the activity must take place at the boundaries of the regions. It turns out that, depending on the relative parities at the boundaries of regions, the boundary will move left one position per step, move right one position per step, or remain in the same place. In the example below, we show a section of the pattern for $n = 175$.

```
9 1 9 1 9 1 9 1 9 1 9 1 9 1 8 0 8 0 7 1 7 1 7 1 6 0 6 0 6 0 6 0 5 0 4 0 4 0 4 0 4 1 3 0 2 0 2 0 2 0 2
1 9 1 9 1 9 1 9 1 9 1 9 1 9 0 8 0 7 1 7 1 7 1 7 0 6 0 6 0 6 0 5 1 4 0 4 0 4 0 4 0 4 1 2 0 2 0 2 0 2 0
9 1 9 1 9 1 9 1 9 1 9 1 9 1 8 0 7 1 7 1 7 1 7 1 6 0 6 0 6 0 5 1 5 0 4 0 4 0 4 0 4 0 4 0 2 0 2 0 2 0 2
1 9 1 9 1 9 1 9 1 9 1 9 1 9 0 7 1 7 1 7 1 7 1 7 0 6 0 6 0 5 1 5 1 4 0 4 0 4 0 4 0 4 0 3 0 2 0 2 0 2 0
9 1 9 1 9 1 9 1 9 1 9 1 9 1 7 1 7 1 7 1 7 1 7 1 6 0 6 0 5 1 5 1 5 0 4 0 4 0 4 0 4 0 3 1 2 0 2 0 2 0 2
1 9 1 9 1 9 1 9 1 9 1 9 1 8 1 7 1 7 1 7 1 7 1 7 0 6 0 5 1 5 1 5 1 4 0 4 0 4 0 4 0 3 1 3 0 2 0 2 0 2 0
9 1 9 1 9 1 9 1 9 1 9 1 9 0 8 1 7 1 7 1 7 1 7 1 6 0 5 1 5 1 5 1 5 0 4 0 4 0 4 0 3 1 3 1 2 0 2 0 2 0 1
1 9 1 9 1 9 1 9 1 9 1 9 1 8 0 8 1 7 1 7 1 7 1 7 0 5 1 5 1 5 1 5 1 4 0 4 0 4 0 3 1 3 1 3 0 2 0 2 0 1 1
9 1 9 1 9 1 9 1 9 1 9 1 9 0 8 0 8 1 7 1 7 1 7 1 5 1 5 1 5 1 5 1 5 0 4 0 4 0 3 1 3 1 3 1 2 0 2 0 1 1 1
1 9 1 9 1 9 1 9 1 9 1 9 1 8 0 8 0 8 1 7 1 7 1 6 1 5 1 5 1 5 1 5 1 4 0 4 0 3 1 3 1 3 1 3 0 2 0 1 1 1 1
9 1 9 1 9 1 9 1 9 1 9 1 9 0 8 0 8 0 8 1 7 1 7 0 6 1 5 1 5 1 5 1 5 0 4 0 3 1 3 1 3 1 3 1 2 0 1 1 1 1 1
1 9 1 9 1 9 1 9 1 9 1 9 1 8 0 8 0 8 0 8 1 7 1 6 0 6 1 5 1 5 1 5 1 4 0 3 1 3 1 3 1 3 1 3 0 1 1 1 1 1 1
9 1 9 1 9 1 9 1 9 1 9 1 9 0 8 0 8 0 8 0 8 1 7 0 6 0 6 1 5 1 5 1 5 0 3 1 3 1 3 1 3 1 3 1 1 1 1 1 1 1 1
1 9 1 9 1 9 1 9 1 9 1 9 1 8 0 8 0 8 0 8 0 8 1 6 0 6 0 6 1 5 1 5 1 3 1 3 1 3 1 3 1 3 1 2 1 1 1 1 1 1 1
9 1 9 1 9 1 9 1 9 1 9 1 9 0 8 0 8 0 8 0 8 0 8 0 6 0 6 0 6 1 5 1 4 1 3 1 3 1 3 1 3 1 3 0 2 1 1 1 1 1 1
1 9 1 9 1 9 1 9 1 9 1 9 1 8 0 8 0 8 0 8 0 8 0 7 0 6 0 6 0 6 1 5 0 4 1 3 1 3 1 3 1 3 1 2 0 2 1 1 1 1 1
9 1 9 1 9 1 9 1 9 1 9 1 9 0 8 0 8 0 8 0 8 0 7 1 6 0 6 0 6 0 6 1 4 0 4 1 3 1 3 1 3 1 3 0 2 0 2 1 1 1 1
1 9 1 9 1 9 1 9 1 9 1 9 1 8 0 8 0 8 0 8 0 7 1 7 0 6 0 6 0 6 0 6 0 4 0 4 1 3 1 3 1 3 1 2 0 2 0 2 1 1 1
9 1 9 1 9 1 9 1 9 1 9 1 9 0 8 0 8 0 8 0 7 1 7 1 6 0 6 0 6 0 6 0 5 0 4 0 4 1 3 1 3 1 3 0 2 0 2 0 2 1 1
1 9 1 9 1 9 1 9 1 9 1 9 1 8 0 8 0 8 0 7 1 7 1 7 0 6 0 6 0 6 0 5 1 4 0 4 0 4 1 3 1 3 1 2 0 2 0 2 0 2 1
9 1 9 1 9 1 9 1 9 1 9 1 9 0 8 0 8 0 7 1 7 1 7 1 6 0 6 0 6 0 5 1 5 0 4 0 4 0 4 1 3 1 3 0 2 0 2 0 2 0 2
1 9 1 9 1 9 1 9 1 9 1 9 1 8 0 8 0 7 1 7 1 7 1 7 0 6 0 6 0 5 1 5 1 4 0 4 0 4 0 4 1 3 1 2 0 2 0 2 0 2 0
9 1 9 1 9 1 9 1 9 1 9 1 9 0 8 0 7 1 7 1 7 1 7 1 6 0 6 0 5 1 5 1 5 0 4 0 4 0 4 0 4 1 3 0 2 0 2 0 2 0 2
1 9 1 9 1 9 1 9 1 9 1 9 1 8 0 7 1 7 1 7 1 7 1 7 0 6 0 5 1 5 1 5 1 4 0 4 0 4 0 4 0 4 1 2 0 2 0 2 0 2 0
9 1 9 1 9 1 9 1 9 1 9 1 9 0 7 1 7 1 7 1 7 1 7 1 6 0 5 1 5 1 5 1 5 0 4 0 4 0 4 0 4 0 4 0 2 0 2 0 2 0 2
1 9 1 9 1 9 1 9 1 9 1 9 1 7 1 7 1 7 1 7 1 7 1 7 0 5 1 5 1 5 1 5 1 4 0 4 0 4 0 4 0 4 0 3 0 2 0 2 0 2 0
9 1 9 1 9 1 9 1 9 1 9 1 8 1 7 1 7 1 7 1 7 1 7 1 5 1 5 1 5 1 5 1 5 0 4 0 4 0 4 0 4 0 3 1 2 0 2 0 2 0 2
1 9 1 9 1 9 1 9 1 9 1 9 0 8 1 7 1 7 1 7 1 7 1 6 1 5 1 5 1 5 1 5 1 4 0 4 0 4 0 4 0 3 1 3 0 2 0 2 0 2 0
9 1 9 1 9 1 9 1 9 1 9 1 8 0 8 1 7 1 7 1 7 1 7 0 6 1 5 1 5 1 5 1 5 0 4 0 4 0 4 0 3 1 3 1 2 0 2 0 2 0 2
1 9 1 9 1 9 1 9 1 9 1 9 0 8 0 8 1 7 1 7 1 7 1 6 0 6 1 5 1 5 1 5 1 4 0 4 0 4 0 3 1 3 1 3 0 2 0 2 0 2 0
9 1 9 1 9 1 9 1 9 1 9 1 8 0 8 0 8 1 7 1 7 1 7 0 6 0 6 1 5 1 5 1 5 0 4 0 4 0 3 1 3 1 3 1 2 0 2 0 2 0 2
```

The moving boundaries produce a nice pattern of triangles, as in the illustration. To explain this pattern we need to know the exact behavior of all boundaries between regions. It is easy enough to check all cases, as there are not many. We

discover that the boundary between two regions having active piles of size $2k + 1$ and size $2k$ will be stationary, and the boundary between regions with piles of size $2k$ and size $2k - 1$ will move toward or away from the origin, depending on whether the inactive pile between the regions contains zero or one disks, respectively. When a moving boundary hits a stationary boundary, it is reflected and the stationary boundary moves over one pile. For example, here is a moving boundary between $2k$ and $2k - 1$ colliding with a stationary boundary between $2k + 1$ and $2k$. The array on the left shows a sample configuration for $k = 4$, and the array on the right shows the boundary being moved.

$$
\begin{array}{ccccccccc}
9 & 1 & 9 & 1 & 8 & 0 & 8 & 0 & 7 \\
1 & 9 & 1 & 9 & 0 & 8 & 0 & 7 & 1 \\
9 & 1 & 9 & 1 & 8 & 0 & 7 & 1 & 7 \\
1 & 9 & 1 & 9 & 0 & 7 & 1 & 7 & 1 \\
9 & 1 & 9 & 1 & 7 & 1 & 7 & 1 & 7 \\
1 & 9 & 1 & 8 & 1 & 7 & 1 & 7 & 1 \\
9 & 1 & 9 & 0 & 8 & 1 & 7 & 1 & 7 \\
1 & 9 & 1 & 8 & 0 & 8 & 1 & 7 & 1 \\
9 & 1 & 9 & 0 & 8 & 0 & 8 & 1 & 7
\end{array}
\qquad\qquad
\begin{array}{cccccccccc}
- & - & - & - & | & - & - & - & < & - \\
- & - & - & - & | & - & - & < & - & - \\
- & - & - & - & | & - & < & - & - & - \\
- & - & - & - & | & < & - & - & - & - \\
- & - & - & | & > & - & - & - & - & - \\
- & - & - & | & - & > & - & - & - & - \\
- & - & - & | & - & - & > & - & - & - \\
- & - & - & | & - & - & - & > & - & - \\
- & - & - & | & - & - & - & - & > & -
\end{array}
$$

We can interpret this pattern in a "physical" way by thinking of the boundaries as "balls" and "walls." Consider a moving boundary to be a ball and a stationary boundary to be a wall. Then, a ball will always be moving one position per step. When a ball hits a wall, the wall recoils one position, and the ball changes direction and bounces off the wall. It is easy to check that when two balls hit a wall simultaneously, the right thing happens: the balls both change direction and the wall does not move. If you break the patterns into regions between walls, the center region is the only region that does not contain a ball. This provides one more rule to determine the game completely: when two walls meet in the center, they both disappear (as do the balls in the adjacent regions).

We now have a certain number of walls with balls bouncing back and forth between them. If the distance between two adjacent walls is $d$, then the ball between these two walls will hit each wall every $2d$ steps, exerting a "pressure" moving the wall one position every $2d$ steps. We can approximate this by a differential equation saying that the wall moves at a rate of $1/2d$.

Each wall is acted on by either one or two balls. The walls at the ends and at the middle only have one ball acting on them, and the other walls have two balls pushing them. We need only look at one side of the origin, say, the right (positive) side. Suppose we have $k$ walls, and we let the position of the $i$th wall from the center be $w_i$. Then we get the differential equations:

$$
\begin{aligned}
\frac{dw_1}{dt} &= \frac{-1}{2(w_2 - w_1)}, \\
\frac{dw_i}{dt} &= \frac{1}{2(w_i - w_{i-1})} - \frac{1}{2(w_{i+1} - w_i)}, \qquad 1 < i < k, \\
\frac{dw_k}{dt} &= \frac{1}{2(w_k - w_{k-1})}.
\end{aligned}
$$

The behavior of the game is approximated by the behavior of this system of differential equations until the middle wall ($w_1$) hits the origin and vanishes. We then replace these differential equations by the corresponding set of equations involving only $k - 1$ walls.

This system of differential equations is invariant under scaling in a very useful way. If we expand each of the regions by a factor of $\alpha$ and decrease the time by a factor of $\alpha^2$, so $w'_k = \alpha w_k$ and $t' = t/\alpha^2$, then a solution of the system of equations is transformed into another solution. This seems to apply to our game. We would like to say that if we start with $\alpha$ times as many disks the solution will be scaled in this manner, with $\alpha$ times as many disks spread over $\alpha$ times the area, and taking $\alpha^2$ times as long. We cannot say this immediately because of two problems: there is a singularity at the starting position, i.e., all the "walls" occupy the origin at $t = 0$, and the "balls hitting walls" pattern does not appear until time $O(n)$. We can overcome these difficulties by starting at a different position and proving a lemma that the change in starting position does not matter.

We can look at the game as an approximation to the differential equation. If we choose some fixed starting position for the differential equation, and let $n$ go to infinity, approximating the differential equation by our scaled disk game, then the behavior of the game will converge to the behavior of the differential equation. When we scale up by $\alpha = 2$, for example, then the ball hits the wall twice as often, and the discrete amount the wall moves becomes half as large. This corresponds to approximating the differential equation by twice as many steps, each half the size.

We will now need a lemma that is essentially a triangle inequality. This lemmas says that if we can reach position $s'$ from position $s$ in $m$ moves, where in a move we are permitted to move any subset of the disks provided that from each pile we move the same number to the right and left, then the number of moves it takes to finish from position $s$ and position $s'$ differ by at most $m$. Thus, if we make $m$ small, the number of moves it takes to finish from $s$ and $s'$ is approximately the same. We denote the number of moves that it takes to finish starting from position $s$ by $M(s)$.

LEMMA 7. *If $s'$ can be reached from $s$ in $m$ moves, where a move involves picking nonnegative integers $r_i$ for every pile $i$ and moving $2r_i$ disks from pile $i$, $r_i$ to pile $i - 1$ and $r_i$ to pile $i + 1$, then $M(s') \leqslant M(s) \leqslant M(s') + m$.*

*Proof.* We will prove each part of the inequality separately. Both proofs are by induction and similar to the proof of Theorem 1.

By Theorem 1, this game where you do not have to move all the disks will always end in the same final position. We first show that in this game, the fastest possible way to finish is to move all the disks possible on every move, i.e., to move as in our original game. Thus, $M(s') + m \geqslant M(s)$, since we can finish from position $s$ in $M(s') + m$ moves by first taking $m$ moves to reach $s'$. This will prove the second part of our inequality.

Suppose that on the first step we do not move all the disks that can be moved. Specifically, suppose that we do not move all possible disks in some pile $i$. Since we must finish with at most 1 disk in each pile, then eventually more disks must be moved off of pile $i$. Suppose this first happens on the $j$th step. Without changing any other moves, we can move two fewer disks off pile $i$ on the $j$th step, and two more on the first step. Continuing this process of pushing moves to the first step,

eventually we will move all possible disks on the first step, and take at most the same number of steps. By induction, we see that moving as many disks as possible on each step will finish as quickly as any other procedure.

We must now prove that $M(s) \geqslant M(s')$. We will show that if $M(s) < M(s')$, then there is a way to finish in $M(s)$ moves starting with $s'$, a contradiction.

Let the position after the first move from $s$ in the original game (moving all disks possible) be $s_1$. We will obtain $s_1'$ from $s'$ by moving everything moved to get from $s$ to $s_1$ that was not moved in getting from $s$ to $s'$. We will then show that we can get from $s_1$ to $s_1'$ in $m-1$ moves, which will prove the theorem by induction.

Let the total number of disks moved off of pile $i$ while going from $s$ to $s'$ be $2d_i$, and let the number of disks on pile $i$ in $s$ be $p_i$. To obtain $s_1'$, move $\max(2\lfloor p_i/2\rfloor - 2d_i, 0)$ disks off pile $i$ in $s'$. There are clearly this many disks on pile $i$ in $s'$, since we started with $p_i$ and removed $2d_i$.

We need to show that we can reach $s_1'$ from $s_1$. We will first show that we can assume that we move $\max(2\lfloor p_i/2\rfloor, 2d_i)$ disks off pile $i$ on the first move going from $s$ to $s'$. We do this by pushing moves to earlier steps. If we move disks off a pile at step $j$ which did not have all possible disks moved on step $j-1$, then we can move two more disks off it on step $j-1$ and two fewer on step $j$ without changing anything else. Repeating this process, we will eventually terminate, since we only push moves to earlier steps. When we get stuck, then for each $i$ either we move all possible disks off pile $i$ on the first step, or we do not move any additional disks off pile $i$ after the first step while going from $s$ to $s'$. The first case happens when $\lfloor p_i/2\rfloor \leqslant d_i$ and the second when $\lfloor p_i/2\rfloor \geqslant d_i$.

It is now easy to see that we can reach $s_1'$ from $s_1$ by using almost exactly the same procedure as we did to get from $s$ to $s'$. If there is more than one unmoved disk in pile $i$ on the first move from $s$ to $s'$, we move these on the first move and then leave them untouched until we reach $s_1'$. This gives a $m$-step path from $s$ to $s_1'$ which moves $2\lfloor p_i/2\rfloor$ disks off pile $i$ on the first move, meaning that after the first step, we reach $s_1$. Since $M(s_1) = M(s) - 1$, and since the lemma is trivial if $M(s) = 1$, this proves Lemma 7 by induction. ∎

We can now use this lemma to prove that there is a constant $c$ such that the number of moves the game takes to finish is $cn^2 + o(n^2)$. We have already shown that $1/3 \leqslant c \leqslant \pi^2/6 - 1$. By using numerical techniques to solve the differential equation, one could approximate the constant arbitrarily closely.

THEOREM 4. *There is a constant c such that* $\lim_{n\to\infty} M(n)/n^2 = c$, *where* $M(n)$ *is the number of moves needed to finish when the starting position is a pile of* $2n+1$ *disks.*

*Proof.* We use the fact that our game approximates the differential equation. Consider the behavior of the differential equation when it starts with walls at positions $w = (w_1, w_2, \dots, w_k)$. The differential equation will take some definite time $c_w$ to finish starting at $w$. We can approximate this with our game if we start with walls at positions $[\alpha w] = (\pm[\alpha w_1], \pm[\alpha w_2], \dots, \pm[\alpha w_k])$. Here, $\alpha$ is any large real number, and $[\alpha w_i]$ is $\alpha w_i$ rounded to either of the adjacent integers. We allow rounding in any consistent way. We will let $s_{\alpha w}$ denote this starting position with walls at $[\alpha w]$. Using the scaling property of the differential equation, we see that $\alpha$ does not affect the behavior of the differential equation, i.e., the differential

equation starting with $\alpha w$ has the same behavior, scaled, as if it started with $w$. Furthermore, if we let $\alpha$ go to infinity in the game and look at it as an approximation to the differential equation starting at $w$, the balls hit the walls more often and the walls move less each time a ball hits. Thus, the behavior of the game goes to the behavior of the differential equation, and $(1/\alpha^2)M(s_{\alpha w}) \to c_w$.

We now need a relationship between $n$ and $\alpha$. The number of disks in $s_{\alpha w}$ is easy to compute. Except possibly for one pair of adjacent piles (the location of the "ball"), the region between $[\alpha w_i]$ and $[\alpha w_{i+1}]$ contains $2(k-i)$ disks in every pair of adjacent piles. Thus, the total number of disks on one side of the origin is

$$k \cdot ([\alpha w_1] - 0) + (k-1) \cdot ([\alpha w_2] - [\alpha w_1]) + \cdots + 1 \cdot ([\alpha w_k] - [\alpha w_{k-1}]) + O(1) = \sum_{i=1}^{k} [\alpha w_i] + O(1),$$

where the constant in the big $O$ notation may involve $k$ and $w$, but not $\alpha$. Thus, the appropriate $n$ for $s_{\alpha w}$ is $n = \sum_1^k [\alpha w_i]$.

We will now apply Lemma 7. To do this, we must show that starting with the appropriate number of disks at the origin, we can reach $s_{\alpha w}$ quickly using the game described in Lemma 7, where it is not required to move all possible disks. If we start with $2n+1 = \sum_1^k (2[\alpha w_i]+1)$ disks, and for each $i$ separately run the original game starting with $2[\alpha w_i]+1$ disks at the origin, then we will end up with $k-i$ disks in every pile between $[\alpha w_i]$ and $[\alpha w_{i+1}]$. If we then move all possible disks in the piles on even integers, we obtain the position $s_{\alpha w}$, with the balls located at walls separating $2j-1$ and $2j$, as in the example below:

| | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 |
| 1 | 1 | 1 | 1 | 2 | 0 | 4 | 0 | 4 | 0 | 5 | 1 | 5 | 1 | 5 | 0 | 4 | 0 | 4 | 0 | 2 | 1 | 1 | 1 | 1 |

The number of steps it takes to reach $s_{\alpha w}$ is one more than the maximum time it takes to finish when starting with $2[\alpha w_i]+1$ disks at the origin, for $1 \leqslant i \leqslant k$. By Lemma 5, this is at most $\alpha^2 w_k^2 + \alpha w_k + 1$. We have, by Lemma 7,

$$\frac{1}{\alpha^2}|M(s_{\alpha w}) - M(n)| \leqslant w_k^2 + \frac{w_k}{\alpha} + \frac{1}{\alpha^2}.$$

For $\alpha$ sufficiently large, we get

$$\left|\frac{1}{\alpha^2}M(s_{\alpha w}) - c_w\right| \leqslant \varepsilon,$$

so combining these equations gives

$$\left|\frac{1}{\alpha^2}M(n) - c_w\right| \leqslant w_k^2 + \varepsilon.$$

Recall that $n/\alpha \approx \sum_1^k w_i$, so for sufficiently large $n$,

$$\left|\frac{1}{n^2}M(n) - \frac{c_w}{\left(\sum_1^k w_i\right)^2}\right| \leqslant 2\left(\frac{w_k}{\sum_i^k w_i}\right)^2.$$

We can make $w_k/\sum_1^k w_i$ arbitrarily small by making $k$ large and choosing evenly

spaced $w_i$. Since $n$ does not appear in $c_w$ or in $w_i$, this proves that $M(n)/n^2$ converges to some constant. ■

In fact, one can show that there is some exact behavior of the walls that is approached by the differential equation as $w_k/\Sigma_1^k w_i$ goes to 0. One does this by showing that if you start an approximate game at a position close to the actual starting position with $2n+1$ disks at the origin (i.e., if it can be reached in $m$ steps, as in Lemma 7), then at any time $t$, the position of the approximate game will be close to the position of the actual game (again, it can be reached in $m$ steps). The fact that the behavior of the disk game approaches a limit as $w_k/\Sigma_1^k w_i \to 0$ implies that the behavior of the differential equation also approaches a limit as $w_k/\Sigma_1^k w_i \to 0$. This limit behavior satisfies the differential equation for $t > 0$. At $t = 0$, there is a singularity; the position at $t = 0$ can be thought of as infinitely many walls at the origin. A "big bang" then happens: walls start disappearing and space appears between the remaining walls. Thus, for very small $t$, there are a very large (but finite) number of walls, which are all very close to (but not at) the origin.

**5. Conclusion.** In this article we have established that the number of moves for the game to reach its final position is $cn^2 + o(n^2)$, where $0.33 \leqslant c \leqslant 0.65$. It is possible to prove somewhat tighter bounds with more sophisticated arguments. Another approach to finding the value of $c$ is by computer simulation. The straightforward simulation algorithm which keeps track of each pile at each unit of time has a run time of $\Omega(n^3)$ so it is only suitable for relatively small values of $n$. A better simulation can be based on the physical view of balls bouncing between walls. Using this model, a program can be written which updates the configuration every time a ball encounters a wall. This gives a considerable savings in time and greatly increases the range that can be simulated. We have simulated the game for values of $n$ up to 10,000. For $n = 10{,}000$, the number of steps was 59,730,533, giving a value of $c \approx 0.5973$.

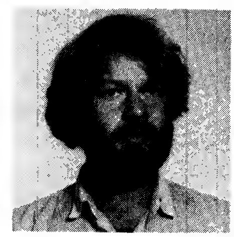
# Mountain Climbing, Ladder Moving, and the Ring-Width of a Polygon

JACOB E. GOODMAN*, *City College, City University of New York*
JÁNOS PACH**, *Hungarian Academy of Sciences*
CHEE K. YAP***, *Courant Institute, New York University*

JACOB ELI GOODMAN received his Ph.D. at Columbia University in 1967, under the direction of Heisuke Hironaka. In addition to algebraic geometry, he has worked in topological graph theory and discrete geometry, and has coauthored a series of papers with Richard Pollack focusing on the interplay between geometric and topological properties of configurations of points and arrangements of hyperplanes and their generalizations.

He is coeditor in chief of *Discrete & Computational Geometry*, a journal of mathematics and computer science published by Springer-Verlag.

JÁNOS PACH received his Ph.D. at Eötvös University, Budapest, in 1980, under the direction of Miklós Simonovits. He is a senior research fellow at the Mathematical Institute of the Hungarian Academy of Sciences, working mainly in combinatorics, discrete geometry, and computational geometry. He is a frequent visitor at the Courant Institute, New York University.



CHEE-KENG YAP was born in Singapore in 1952, grew up in Malaysia, and attended MIT (B.S., 1975) and Yale (Ph.D., 1980). Since 1981, he has been at the Computer Science Department of the Courant Institute of Mathematical Sciences, New York University. He is an editor of *S.I.A.M. Journal of Computing* and the *Journal of Computer and System Sciences*. His research interests are mainly (ultimately) algorithmic, whether in the guise of complexity theory, symbolic and algebraic computation, computational geometry, or robotic 'applications' as in this article.

**Introduction.** In this article we consider several geometric problems which at first glance seem only vaguely connected, but which on closer examination turn out to be intimately related. The first is well known in mathematical folklore (but the proof presented here seems to be new), the second is an amusing variant of the first, and the third is a motion-planning problem arising in the relatively new mathematical discipline of robotics; in fact it was this third problem, suggested in [8], which motivated our investigation. (See [7] for a general survey of the area of motion planning.)

*Supported in part by NSF grant DMS-8501492 and PSC-CUNY grant 666426
**Supported in part by Courant Institute (NYU) and Hungarian NSFR grant 1812
***Supported in part by NSF grants DCR-8401898 and DCR-8401633

(I) Two mountain climbers begin at sea level, at opposite ends of a (two-dimensional) chain of mountains. Can they find routes along which to travel, always maintaining equal altitudes, until eventually they meet?

(II) Two painters are carrying a 20-foot ladder, one at each end, along a garden path which begins and ends with long straight segments. Can they negotiate the path without violating the "keep off the grass" signs? Moreover, what happens if the path splits into two paths which subsequently reunite: can the painters separate and later rejoin each other?

(III) If an irregularly shaped body can be dragged completely (i.e., continuously translated and rotated) through two posts a distance $d$ apart, can it be dragged completely through two posts separated by a distance greater than $d$?

Each of these problems seems to have an intuitively clear solution which, on closer examination, proves decidedly elusive; this is true especially of the second and third. However, by the time we are done, we will have seen not only that the problems are interconnected in that the solutions to the first two are needed for the solution to the third, but that the most important common feature of all of the problems is that once we translate them into the language of elementary plane topology, their solutions become natural and transparent.

First we present a simple topological lemma which lies at the heart of the proofs.

LEMMA 1. *Given a square $S$ with consecutive vertices $A$, $B$, $C$, $D$, and a graph $\Gamma$ (by which we mean here a finite system of [not necessarily simple] closed arcs and intersection points) lying in $S$ and containing $A$ and $C$ and no other point of the boundary of $S$. If $A$ and $C$ are not in the same connected component of $\Gamma$ then $B$ and $D$ can be joined by a path lying in $S$ which avoids $\Gamma$.*

*Proof.* Replace $\Gamma$ by a tubular neighborhood $\overline{\Gamma}$ of sufficiently small radius $\varepsilon$. Then, starting at $A$, run along the boundary of $\overline{\Gamma}$. (The path we follow is well defined, since—even though $\Gamma$ may branch, as in Fig. 1—the boundary of $\overline{\Gamma}$ does not.) Eventually we must either return to $A$ or arrive at $C$, since the total number of arcs is finite and we cannot repeat a previous position because no branching takes place. In the first case the path itself (together with part of the boundary of $S$) constitutes a path from $B$ to $D$ which avoids $\Gamma$, and in the second case the corresponding arcs of $\Gamma$ constitute a path from $A$ to $C$. (See Fig. 1 for illustrations of both possibilities.) □

(Notice that Lemma 1 can be thought of as a kind of weak converse to the Jordan Curve Theorem: From the Jordan theorem it follows immediately that if there is a path joining a pair of opposite corners of a square which lies entirely within the square, then any path joining the two remaining corners and lying in the square must cross the first; Lemma 1 asserts that if there is no path which can be pieced together from a given finite collection of arcs to connect a pair of opposite corners, then the remaining two corners can be connected by a path not touching any of the arcs.)

We shall begin the discussion of each of the three problems by giving it a precise formulation, and in some cases discuss the consequences of choosing our formulations with less care. The article concludes with the derivation of an algorithm which determines the minimum separation $d$ for which the condition in Problem III holds. We thank Walter Daum for providing the example illustrated in Fig. 9, and we note,

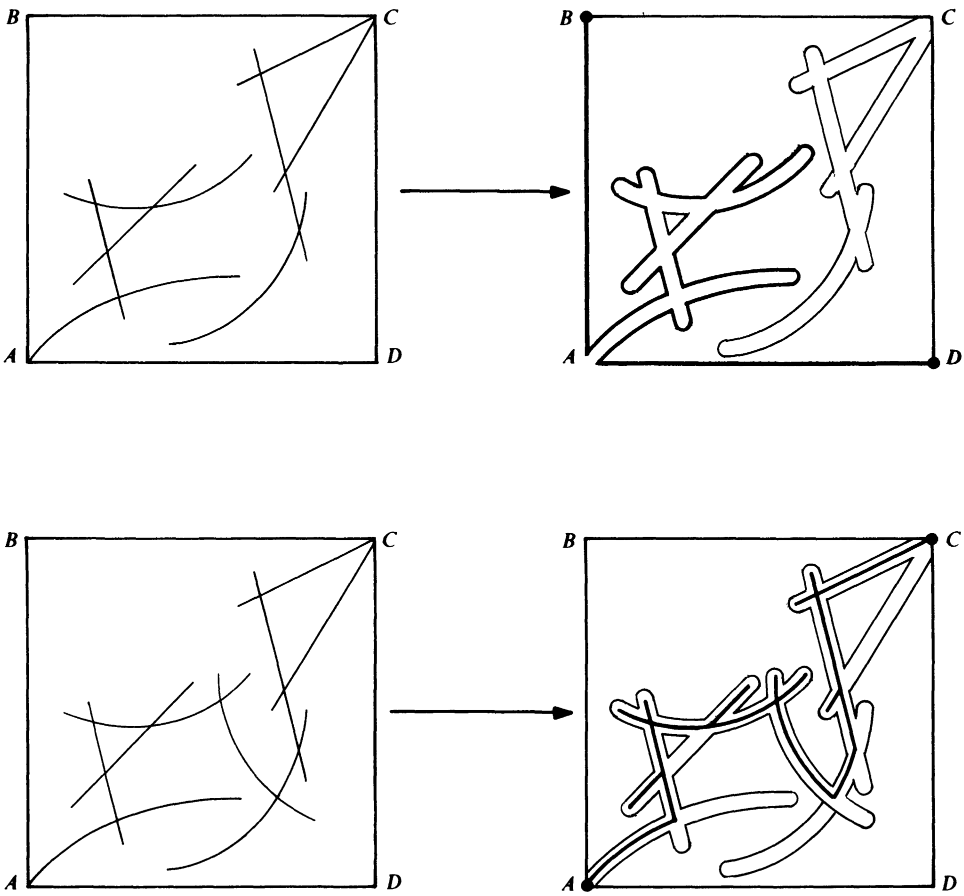



FIG. 1

in addition, that Theorem 2 and some weaker versions of Theorem 4 have also been proved independently by Fan Chung and Paul Seymour [2], David Eppstein [3], and John Kutcher and Joseph O'Rourke [4], respectively.

**1. The mountain climbers' problem.** Our assumption will be that the silhouette of the mountain chain is piecewise monotone. As a matter of fact, we can then also suppose it to be piecewise linear, by replacing each maximal monotone portion by the line segment joining its endpoints. If we now select a point of maximum altitude and reparametrize, it becomes sufficient to prove the following theorem:

THEOREM 1. *Let* $s(\tau)$ *and* $s'(\tau)$ *be continuous, piecewise linear functions from* $[0,1]$ *to* $[0,1]$, *with* $s(0) = s'(0) = 0$ *and* $s(1) = s'(1) = 1$. [*Note*: *the "prime" does not represent differentiation here*!] *Then there exist two other continuous, piecewise linear functions*

$$\tau(x), \tau'(x) : [0,1] \to [0,1]$$

*such that*

$$\begin{aligned} \tau(0) &= \tau'(0) = 0, \\ \tau(1) &= \tau'(1) = 1, \\ s(\tau(x)) &= s'(\tau'(x)) \quad \textit{for every } x \in [0,1]. \end{aligned} \tag{1}$$

*Moreover, $\tau$ and $\tau'$ have a total of $O(n^2)$ critical points, where $n$ denotes the total number of critical points of $s$ and $s'$.*

(In other words, if $A$ and $B$ are the mountain climbers, we have rephrased the problem so that they each start at time 0 and height 0 and want to wind up at time 1 and height 1, maintaining the same height as they move. $A$ and $B$ can backtrack, of course, and their motion is represented by the functions $\tau(x)$ and $\tau'(x)$ which we seek.)

*Proof.* The essence of the argument is to consider the set $\Gamma$ of pairs of positions which are at the same height, and to show that we can get from the bottom (both at height 0) to the top (both at height 1) without ever leaving $\Gamma$. Hence let $\Gamma$ be the set of pairs $(\tau, \tau')$ for which $s(\tau) = s'(\tau')$. Then both $(0,0)$ and $(1,1)$ are contained in $\Gamma$, and we must show that they belong to the same arcwise connected component. There are two ways to proceed. The first, which is simpler, uses Lemma 1 after first observing that nothing changes if we assume that the only points on the silhouettes at height 0 or 1 are the beginning and the end. The simple proof based on these hints is left to the reader. However, in order to gain greater insight into the situation, we will instead do a little extra work, and show that $\Gamma$ is the union of a finite number of isolated points and 1-manifolds with boundary and is made up of at most $n^2$ line segments, and that in fact the only boundary points arising are $(0,0)$ and $(1,1)$ themselves.

To this end, we examine a pair of points $(\tau, s(\tau))$ and $(\tau', s'(\tau'))$ with $s(\tau) = s'(\tau')$, i.e., a pair of positions of $A$ and $B$ which are at the same height. There are several cases, each illustrated in Fig. 2, where arrows with the same number of arrowheads represent corresponding motions of $A$ and $B$ and of the point of $\Gamma$ which represents their joint position.

(a) If $A$ and $B$ are each at interior points of maximal linear segments of $s$ and $s'$, respectively, then they can each move in two directions (either both forward or both backward if the segments have the same monotonicity, or one forward and the other backward if the segments have opposite monotonicity). Hence in the neighborhood of the corresponding point of $\Gamma$, $\Gamma$ is a 1-manifold (in fact a line segment!) (Fig. 2a).

(b) If one of $A$ and $B$, say $A$, is at a local maximum (or minimum) of $s$, and the other is at an interior point of a segment of $s'$, then $A$ can move in two directions and $B$ in one, so again $\Gamma$ is a 1-manifold in a neighborhood of the corresponding point (Fig. 2b).

(c) If $A$ and $B$ are each at local maxima or each at local minima of $s$ (resp. $s'$), then there are precisely four ways for them to proceed maintaining the same height: $A$ and $B$ can each move either forward or backward. Thus the corresponding point of $\Gamma$ looks like a simple crossing (Fig. 2c).

(d) If $A$ is at a local maximum of $s$ and $B$ at a local minimum, or vice versa, there is no way for them to proceed (or, for that matter, to have gotten there in the first place!), so the corresponding point of $\Gamma$ is isolated (Fig. 2d).

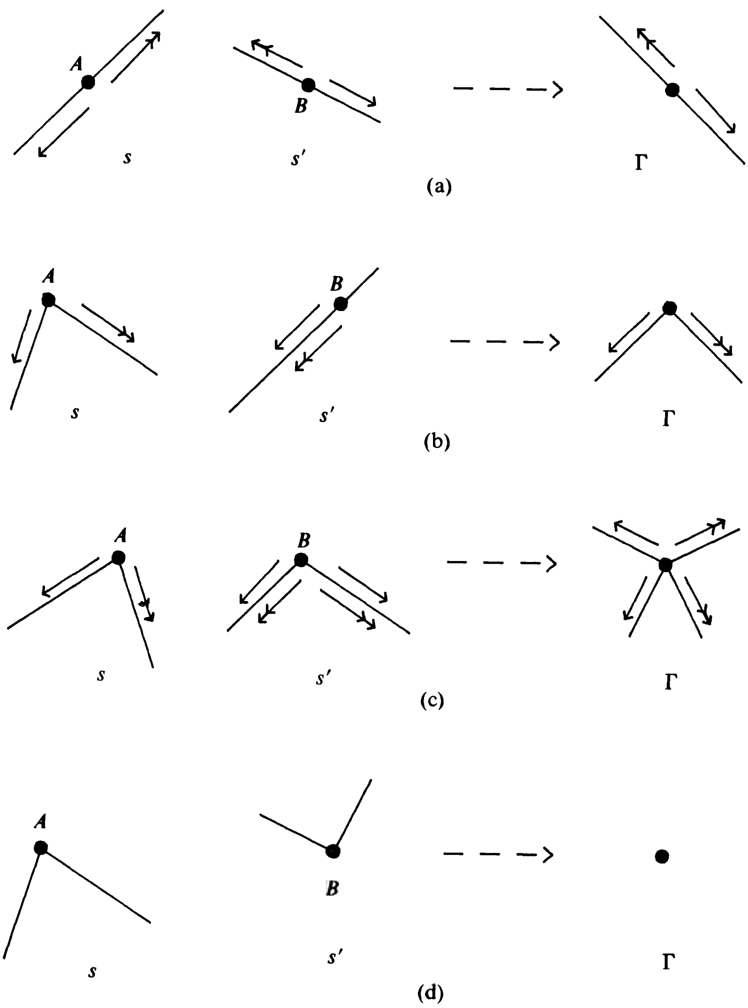



FIG. 2

Now suppose $A$ and $B$ are moving along $s$ and $s'$, respectively. Then they continue their motion until one of them hits a critical point. Since the other is at the same height, and since each of $s$ and $s'$ has $O(n)$ critical points (hence $O(n)$ segments), this can happen in at most $O(n^2)$ ways. The only situations in which there is only one way for $A$ and $B$ to move are when both are at $(0,0)$ or both at $(1,1)$; hence the corresponding points of $\Gamma$ are the only boundary points. Thus $\Gamma$ is the union of $O(n^2)$ line segments and isolated points, and (throwing away the isolated points), each vertex of $\Gamma$ has even degree except for $(0,0)$ and $(1,1)$, which means that these must lie in the same connected component of $\Gamma$. □

Notice that if we allowed the mountain silhouette to dip below sea level, the conclusion would no longer be valid, as is shown by Fig. 3.

We mention also, for completeness, that there is yet another proof, which proceeds by induction on the number of monotone segments in the silhouette. We

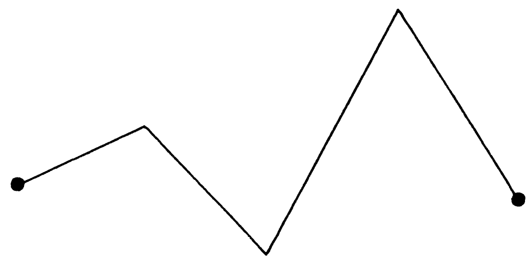


FIG. 3

have chosen to disregard this proof, however, since we wish to focus on a topological approach that applies to all three problems.

**2. The ladder movers' problem.** To answer the first part of Problem II of the Introduction, we prove:

THEOREM 2. *Suppose* $x = (x_1(s), x_2(s))$, $0 \leqslant s \leqslant L$, $L > 1$ *is a polygonal curve, parametrized by its arc length, whose initial and final segments each have length at least* 1, *and which has the property that the only points within unit distance of* $x(0)$ *are the points* $x(s)$ *with* $0 \leqslant s \leqslant 1$, *and the only points within unit distance of* $x(L)$ *are the points* $x(s)$ *with* $L - 1 \leqslant s \leqslant L$. *Then a line segment of unit length can be moved continuously from* $\overline{x(0), x(1)}$ *to* $\overline{x(L-1), x(L)}$ *so that its endpoints* $A$ *and* $B$ *never leave the curve.*

*Proof.* This time we let $\Gamma$ be the set of pairs $(s, t)$ such that $|x(s) - x(t)| = 1$. A line segment $\overline{AB}$ of unit length makes four angles with the curve $x$: two at $A$ (call them $A_1$ and $A_2$), and two of $B$ ($B_1$ and $B_2$). Assume for the moment that no two segments of the polygonal path are parallel and exactly one unit apart. Again examining the several cases that may arise, we see that either $A$ and $B$ can move together in two directions (if exactly one of the $A_i$ and exactly one of the $B_i$ is acute, as in Fig. 4a, the general case), or one can move in two directions and the other in one (if exactly one or three of $A_1$, $A_2$, $B_1$, $B_2$ is acute, as in Fig. 4b), or each can move in either of two directions, giving a total of four possibilities (if both $A_i$ are acute, and neither $B_i$ is, or vice versa, as in Fig. 4c), or neither can move (if all of the $A_i$ and $B_i$ are acute, or none of them is, so that the situation is actually impossible to reach, as in Fig. 4d). There is one additional possibility: it may happen that two segments of the polygonal path $x$ are parallel, and precisely one unit apart. In this case, if $A$ (say) is at a vertex, there will be either one or three ways the motion can proceed: Fig. 4e illustrates the latter possibility. Thus the situation is very much as in the case of Theorem 1, although the curves which comprise $\Gamma$ are now either arcs of (or complete) ellipses or line segments, but with the major difference that the degrees of some vertices of $\Gamma$ may now be odd. But—except for the "starting" and "ending" vertices of $\Gamma$, i.e., the points $(0, 1)$ and $(L - 1, L)$—this happens only when two segments of $x$ are parallel and distance 1 apart, so that the corresponding arc of $\Gamma$ beginning at such a vertex of odd degree must be a line segment which also ends at a vertex of odd degree. Thus we can remove all such offending arcs to obtain a new graph $\Gamma'$, which shares all the topological properties of the graph $\Gamma$ in the proof of Theorem 1, and the conclusion follows. $\square$

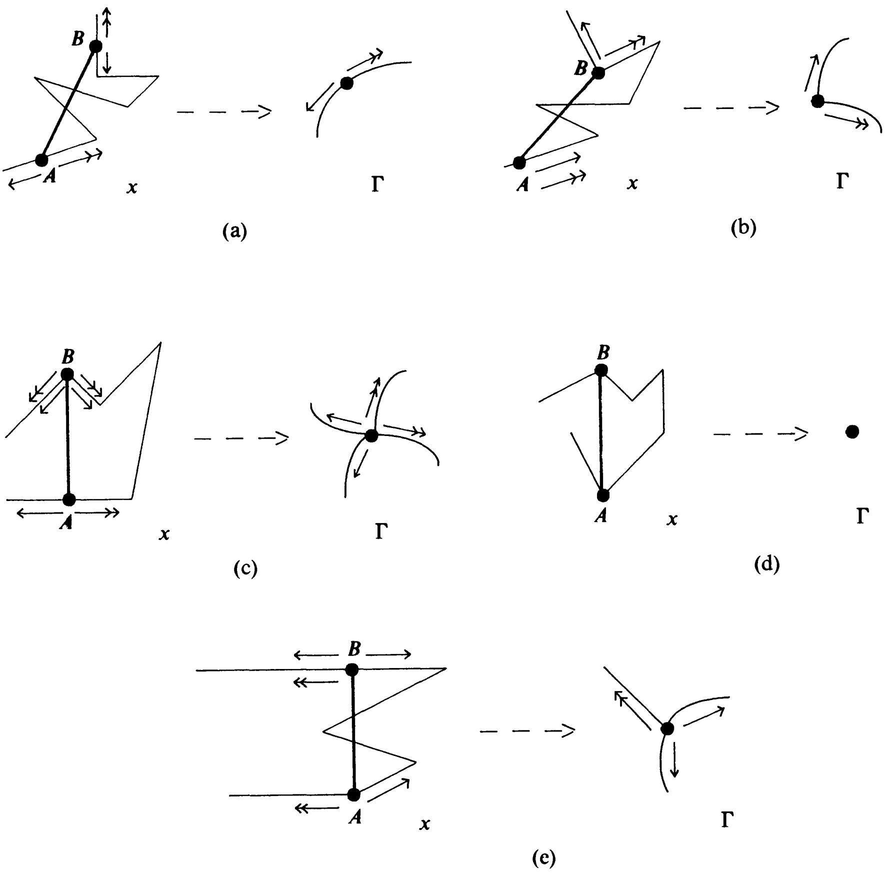



FIG. 4

The second part of Problem II will be addressed by Theorem 3, which may be paraphrased as follows: Suppose the two paths, which we again assume polygonal, diverge at point $P$ and rejoin at point $Q$. If two people, one walking from $P$ to $Q$ along one branch, the other from $Q$ to $P$ along the other branch, cannot move in such a way that they stay more than 20 feet apart all the way, then the painters can indeed carry the 20-foot ladder along the two paths from $P$ to $Q$.

THEOREM 3. *Let $\rho_1, \rho_2$ be two polygonal paths between points $c_1$ and $c_2$, i.e., piecewise linear arcs $\rho_i$: $[0,1] \to \mathbf{R}^2$ with $\rho_1(0) = \rho_2(0) = c_1$, $\rho_1(1) = \rho_2(1) = c_2$. Assume further that*

$$\begin{aligned}
&\rho_1\left(\tfrac{1}{3}\right) = \rho_2\left(\tfrac{1}{3}\right) = c_1^*,\ \rho_1\left(\tfrac{2}{3}\right) = \rho_2\left(\tfrac{2}{3}\right) = c_2^* && \text{for some } c_1^*, c_2^*,\\
&|c_1 - c_1^*| = |c_2 - c_2^*| = d && \text{for some } d > 0,\\
&\rho_1(t) = \rho_2(t) && \text{for every } t \in \left[0, \tfrac{1}{3}\right] \cup \left[\tfrac{2}{3}, 1\right],\\
&|c_1 - \rho_i(t)| > d \text{ and } |c_2 - \rho_i(1-t)| > d && \text{for every } t \in \left(\tfrac{1}{3}, 1\right], i = 1, 2.
\end{aligned} \tag{2}$$

*Then at least one of the following statements is true*:

(i) *Point $A$ can move from $c_1$ to $c_2^*$ along $\rho_1$, and point $B$ can move from $c_1^*$ to $c_2$ along $\rho_2$, so that their distance is always equal to $d$.*

(ii) *$A$ can move from $c_1$ to $c_2$ along $\rho_1$, and $B$ can move from $c_2$ to $c_1$ along $\rho_2$, so that their distance is always larger than $d$.*

*Proof.* (See Fig. 5.) Suppose (ii) does not hold. Let $\Gamma$ represent the pairs of points a unit distance apart; this time

$$\Gamma = \{(s,t) \mid 0 \leqslant s, t \leqslant 1|, |\rho_1(s) - \rho_2(t)| = d\}.$$

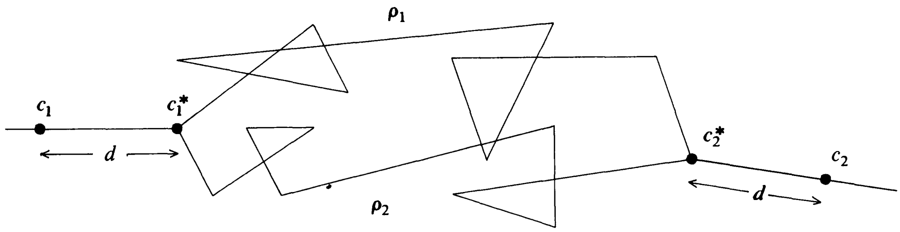



FIG. 5

Again, as in the proof of Theorem 2, $\Gamma$ consists of only finitely many arcs of ellipses, line segments, and isolated points. This time, however, there are four points at which $\Gamma$ meets the boundary of the square $0 \leqslant s \leqslant 1, 0 \leqslant t \leqslant 1$, as Fig. 6 shows. We must show that $(0, 1/3)$ can be joined to $(2/3, 1)$ along $\Gamma$.

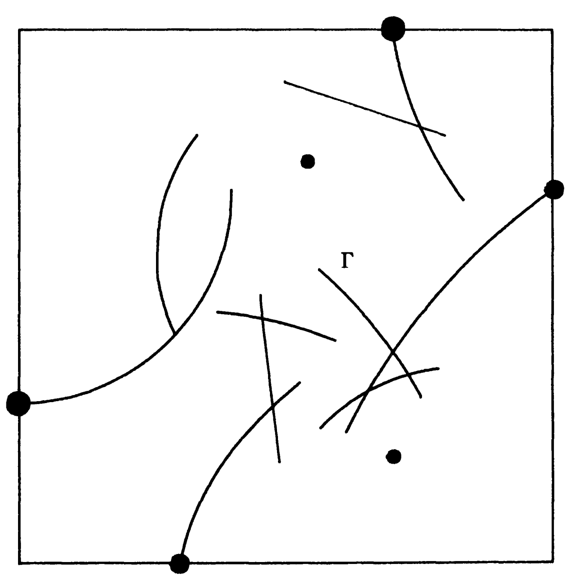



FIG. 6

Considering the possible critical points of the motion of $A$ and $B$, we see that exactly the same cases arise as in the proof of Theorem 2 (it doesn't matter, locally, whether portions of the curves on which the points $A$ and $B$ move are on "parallel" tracks, or not). Hence, as in the proof of Theorem 2, after removing the parts of $\Gamma$ arising from possible motion along parallel line segments (as well as any isolated points), we see that $\Gamma$ contains no vertex of odd degree except for the four points on the boundary of the square. Hence the only cases arising are those shown (schematically) in Fig. 7. In case (a) it follows from Lemma 1 that there is a path from $(0,1)$ to $(1,0)$ which avoids $\Gamma$, and this means that alternative (ii) holds, contrary to our assumption. Thus we must be in either case (b) or case (c), but in each of these cases there is a path in $\Gamma$ with the desired property. □

The condition that the curve in Theorem 2 is polygonal is not superfluous, as is shown by

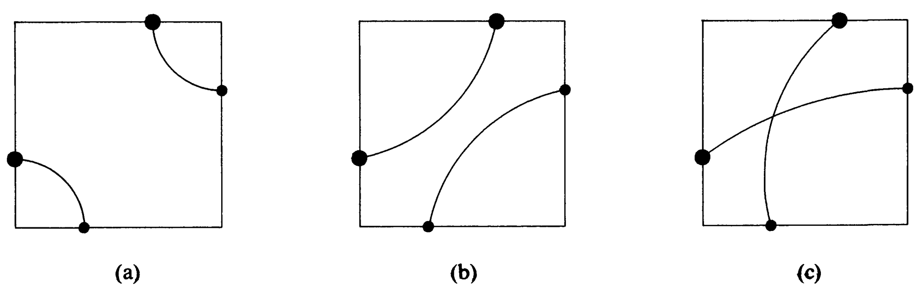



FIG. 7

**PROPOSITION** 1. *There is a simple (i.e., non-self-intersecting) continuous arc $x^*$, which starts and ends with nonparallel half lines $h_1$ and $h_2$, and which admits no continuous motion of a line segment of unit length from $h_1$ to $h_2$ along $x^*$.*

*Proof.* Let $c_p$, $c_q$, and $c_r$ denote the unit circles around points $p$, $q$, and $r$, respectively, with $p, q \in c_r$. Let $h_1$ and $h_2$ be two half lines with endpoints $p$ and $r$, respectively, and let $x^*$ be defined as the union of $h_1$, $h_2$, the portion of $c_r$ between $p$ and $q$, and an arc connecting $q$ and $r$, which meets $c_p$ and $c_q$ infinitely many times, as in Fig. 8.

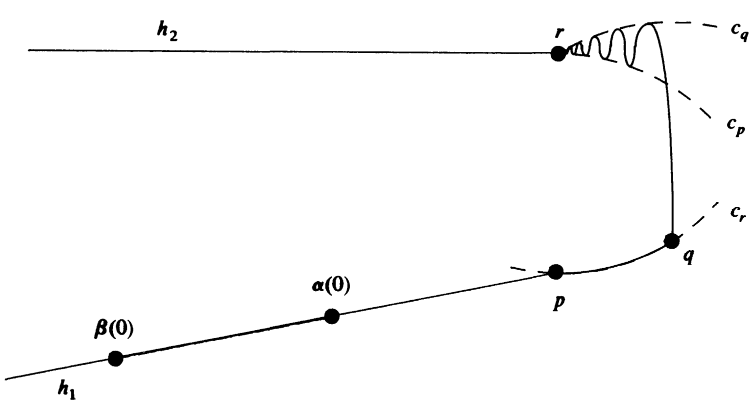



FIG. 8

Assume now, in order to obtain a contradiction, that there are continuous functions $\alpha(t), \beta(t) : [0,1] \to x^*$ such that $|\alpha(t) - \beta(t)| = 1$ for every $t \in [0,1]$ and such that $\alpha(0), \beta(0) \in h_1$ and $\alpha(1), \beta(1) \in h_2$, with $\alpha(0)$ between $\beta(0)$ and $p$. Let

$$t^* = \inf\{t \in [0,1] \mid \alpha(t) = r\}.$$

Then, given any $\varepsilon > 0$, $\alpha(t) \in c_p$ for infinitely many values $t \in (t^*, t^* - \varepsilon)$, and similarly $\alpha(t) \in c_q$ for infinitely many values $t \in (t^*, t^* - \varepsilon)$. But $\beta(t) = p$ whenever $\alpha(t) \in c_p$, and similarly $\beta(t) = q$ whenever $\alpha(t) \in c_q$; hence $\lim_{t \to t^*} \beta(t)$ cannot exist, contradicting our assumption. □

By a slight modification of this construction we can obtain a smooth (i.e., differentiable) simple curve which does not permit a continuous, rectifiable motion of the ladder.

Moreover, the assumptions about $x(0)$ and $x(L)$ in Theorem 2 are essential, as shown by the example illustrated in Fig. 9.

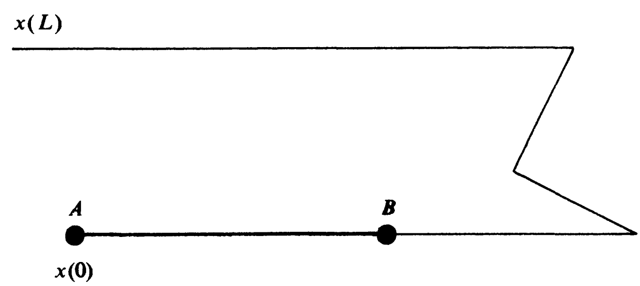



FIG. 9

**3. The ring-width problem.** Let $P$ be a simple closed polygon in the Euclidean plane $\mathbf{R}^2$ with two handles (half lines), $h_1$ and $h_2$, sticking out at two points of its boundary. Assume that these handles do not point in the same direction, and that they do not meet each other or $P$ at any other points (see Fig. 10).

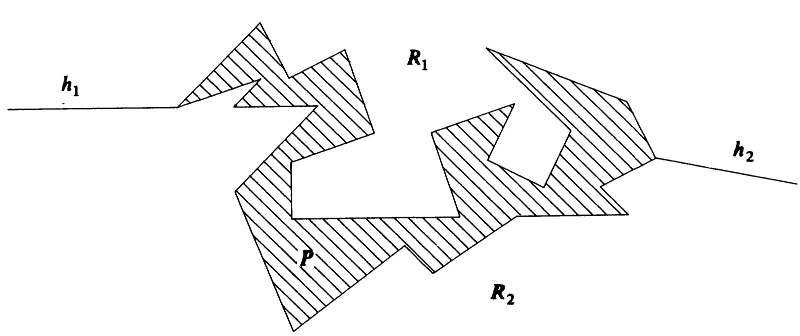



FIG. 10

With some distortion of language (which may be justified by the 3-dimensional analogue of the problem), the *ring-width* $w_R(P)$ of $P$ is defined as the smallest

number $w$ such that a "ring of diameter $w$" can be pulled from one handle of $P$ to the other. To be precise, let $R_1$ and $R_2$ denote the closures of the two unbounded regions of $\mathbf{R}^2 \setminus (h_1 \cup P \cup h_2)$. Then $w_R(P)$ is the least $w$ having the property that there exist two continuous arcs $\pi_1$: $[0,1] \to R_1$, $\pi_2$: $[0,1] \to R_2$ satisfying

$$|\pi_1(t) - \pi_2(t)| = w \quad \text{for every } t \in [0,1], \tag{3}$$

and such that

$$\overline{\pi_1(0), \pi_2(0)} \cap (h_1 \cup P \cup h_2) = \overline{\pi_1(0), \pi_2(0)} \cap h_1 \neq \varnothing$$

and

$$\overline{\pi_1(1), \pi_2(1)} \cap (h_1 \cup P \cup h_2) = \overline{\pi_1(1), \pi_2(1)} \cap h_2 \neq \varnothing .$$

The *elastic ring-width* of $P$, $w_{ER}(P)$, is defined similarly, the only difference being that the ring is now allowed to contract and expand up to size $w$; i.e., instead of (3) only

$$|\pi_1(t) - \pi_2(t)| \leqslant w \quad \text{for every } t \in [0,1] \tag{3'}$$

is required. Obviously $w_{ER}(P) \leqslant w_R(P)$. We will prove the following assertions, the first of which addresses Problem III of the Introduction:

THEOREM 4. *Any ring of diameter $d \geqslant w_R(P)$ can be pulled from one handle of $P$ to the other.*

THEOREM 5. $w_R(P) = w_{ER}(P)$ *for any $P$.*

(Warning: It takes some time to realize that neither of the two statements above is trivial! The notions of ring-width and elastic ring-width were introduced in [**8**] and both of these results were conjectured, but for polygons without handles; that case still remains open.)

Let $B_i$ denote the boundary of $R_i$ ($i = 1,2$). A pair of arcs $(\pi_1, \pi_2)$ with the properties above, including (3) rather than (3′), is called *a contact motion of a ring* of diameter $w$ along $P$ if, in addition, $\pi_i$: $[0,1] \to B_i$ ($i = 1,2$). *A contact motion of an elastic ring* is defined similarly, except that now condition (3) should be replaced by (3′).

THEOREM 6. *There is a contact motion of a ring of diameter $d$ along $P$ if and only if there is a contact motion of an elastic ring of diameter $d$ along $P$.*

Theorems 4–6 are easy consequences of:

THEOREM 7. *Every polygon $P$ with handles permits a contact motion of a ring of diameter $d$, for any $d \geqslant w_{ER}(P)$.*

Before we can prove Theorem 7, we need one more lemma, which can be paraphrased as follows: If $A$ and $B$ are walking from city $c_1$ to city $c_2$ along roads $\gamma_1$ and $\gamma_2$, respectively, and they hold a rope stretched between them, then at some moment they will necessarily catch $C$ who wants to get from $c_2$ to $c_1$ at the same time within the region enclosed by $\gamma_1$ and $\gamma_2$.

LEMMA 2. *Let $\gamma : [0,1] \to \mathbf{R}^2$ be a simple arc connecting two points $c_2$ and $c_1$, with $\gamma(0) = c_2$ and $\gamma(1) = c_1$, and let $h_1$ and $h_2$ be two disjoint half lines whose endpoints are $c_1$ and $c_2$, respectively. Assume further that no interior point of $\gamma$ is on $h_i$, and let*

*$R_1$ and $R_2$ denote the closures of the connected components of $\mathbf{R}^2 \setminus (h_1 \cup h_2 \cup \gamma)$. Let $\gamma_i : [0,1] \to R_i$ $(i = 1,2)$ be (possibly self-intersecting) arcs with $\gamma_i(0) = c_1$, $\gamma_i(1) = c_2$.*

*Then there exists $x \in [0,1]$ such that $\gamma(x)$ lies on the line segment $\overline{\gamma_1(x), \gamma_2(x)}$.*

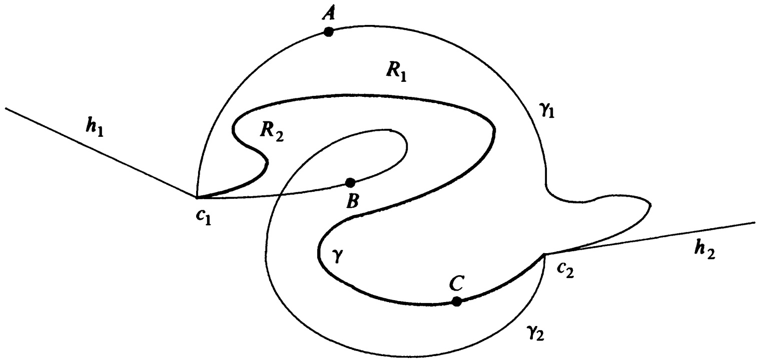



FIG. 11

*Proof.* (See Fig. 11.) For any $t \in [0,1]$, let us denote $\gamma_1(t)$, $\gamma_2(t)$, $\gamma(t)$ by $A$, $B$, $C$, respectively. If, during the motion, $C$ ever coincides with either $A$ or $B$, we are done. So suppose not. As soon as $A$, $B$, and $C$ are under way, i.e., at time $t$ with $0 < t < 1$, consider the directed arc consisting of $h_1$ (reversed), $\gamma_1$ from $c_1$ to $A$, the directed line segment $\overrightarrow{AB}$, $\gamma_2$ from $B$ to $c_2$, and $h_2$. For $t$ close to 0, point $C$ is in the region "above" this directed path, while for $t$ close to 1, $C$ is "below" it (see Fig. 12). Since everything varies continuously with $t$, $C$ must cross the path somewhere. But it never does so outside of the segment $\overline{AB}$, hence we are done.

(Notice that what we are really talking about is the winding number about $C$ of the closed curve $\Gamma$ consisting of the directed path described above plus an arc of a large circle going counterclockwise from a point far out on $h_2$ to a point far out on

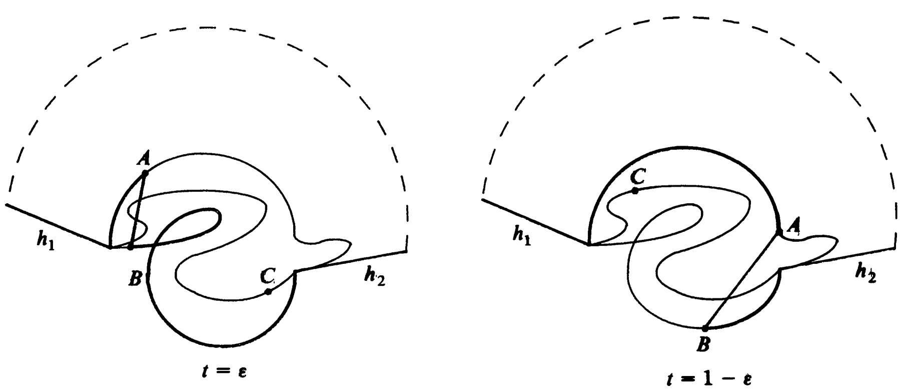



FIG. 12

$h_1$. Since this winding number,

$$\frac{1}{2\pi i}\int_\Gamma \frac{dz}{z-\gamma(t)},$$

as long as it is defined, varies continuously with $t$ as $t$ goes from $\varepsilon$ to $1-\varepsilon$ and changes from 1 to 0, remaining locally constant (since it is an integer!), it must fail to be defined for some value of $t$ between 0 and 1, which implies the result.)

*Proof of Theorem 7.* Let us fix $d \geqslant w_{ER}(P)$, and consider the motion of an elastic ring of diameter $w_{ER}(P) = w$ along $P$, i.e., a pair of continuous maps $\pi_1$: $[0,1] \to R_1$, $\pi_2$: $[0,1] \to R_2$, such that (3′) holds. (See Fig. 10.) Assume without loss of generality that the graph of $\pi_i$ is a polygonal path $\bar{\rho}_i$ satisfying conditions (2) for some $c_i, c_i^* \in h_i$ $(i = 1,2)$. That is,

$$\pi_1(\tau) = \bar{\rho}_1(s(\tau)), \qquad \pi_2(\tau) = \bar{\rho}_2(t(\tau)) \quad \text{for all } \tau \in [0,1], \tag{4}$$

for some continuous, piecewise monotone functions $s(\tau)$, $t(\tau)$: $[0,1] \to [0,1]$ with

$$s(0) = t(0) = 0, s(1) = t(1) = 1. \tag{5}$$

Further, Let $\beta_i$: $[0,1] \to B_i$ be a piecewise linear "parametrization" of the portion of $B_i$ between $c_1$ and $c_2$ such that

$$\beta_i(\tau) = \bar{\rho}_1(\tau) = \bar{\rho}_2(\tau) \quad \text{for every } \tau \in \left[0,\tfrac{1}{3}\right] \cup \left[\tfrac{2}{3},1\right], \qquad i = 1,2.$$

First we are going to show that a person $A$ can walk from $c_1$ to $c_2^*$ along $\bar{\rho}_1$, and $B$ can walk from $c_1^*$ to $c_2$ along $\beta_2$, so that their distance is always equal to $d$. Roughly speaking, this means that there is a motion of a ring of diameter $d$ around $P$, which is a "contact motion from below."

Assume that this is not true. Then we can apply Theorem 3 with $\rho_1 = \bar{\rho}_1$, $\rho_2 = \beta_2$ to conclude that $A$ can walk from $c_1$ to $c_2$ along $\bar{\rho}_1$, and $B$ can walk from $c_2$ to $c_1$ along $\beta_2$, so that their distance is always larger than $d$. That is,

$$|\bar{\rho}_1(s'(\tau)) - \beta_2(t'(\tau))| > d \qquad \text{for all} \tau \in [0,1] \tag{6}$$

for some continuous, piecewise monotone functions $s'(\tau)$, $t'(\tau)$: $[0,1] \to [0,1]$ with

$$s'(0) = t'(1) = 0, s'(1) = t'(0) = 1. \tag{7}$$

By (5) and (7), we can apply Theorem 1 to obtain the fact that there exist two continuous, piecewise monotone functions $\tau(x)$, $\tau'(x)$: $[0,1] \to [0,1]$ satisfying (1). Then (4) and (3′) imply that

$$|\bar{\rho}_1(s(\tau(x))) - \bar{\rho}_2(t(\tau(x)))| = |\pi_1(\tau(x)) - \pi_2(\tau(x))| \leqslant w \leqslant d.$$

On the other hand, by (6) and (1),

$$|\bar{\rho}_1(s(\tau(x))) - \beta_2(t'(\tau'(x)))| = |\bar{\rho}_1(s'(\tau'(x))) - \beta_2(t'(\tau'(x)))| > d.$$

Applying Lemma 2 to the three curves $\gamma_1(x) = \bar{\rho}_1(s(\tau(x)))$, $\gamma_2(x) = \bar{\rho}_2(t(\tau(x)))$, and $\gamma(x) = \beta_2(t'(\tau'(x)))$, $x \in [0,1]$, we obtain the fact that $\gamma(x)$ lies on the segment $\overline{\gamma_1(x), \gamma_2(x)}$ for some $x \in [0,1]$. This contradicts the last two inequalities.

Thus, we can conclude that there is a motion of a ring of diameter $d$ along $P$, which is a "contact motion from below." Performing the same trick again (with $\beta_1$, instead of $\beta_2$), we obtain a motion of a ring of diameter $d$ along $P$, which is a contact motion (both from below and from above). This completes the proof. □

**4. Computing the ring-width.** Since our original motivation arose from the practical subject of robotics, let us now address the question of computation. We shall exploit the existence of contact motions (Theorem 7) to derive an algorithm for determining the ring-width of $P$. If $P$ has $n$ vertices then the algorithm runs in time $O(n^2 \log n)$. (Recall that while $O(f(n))$ means $\leqslant c_1 f(n)$, $\Omega(f(n))$ means $\geqslant c_2 f(n)$ for some constants $c_1, c_2 \neq 0$ and $n$ sufficiently large; we shall use $\Omega(f(n))$ more loosely to mean $\geqslant c_2 f(n)$ for infinitely many values of $n$.) But first it is worthwhile pointing out that time $\Omega(n^2)$ is unavoidable for any algorithm that actually computes a contact motion attaining the minimum width (our algorithm has this property).

Consider the polygonal path shown in Fig. 13.

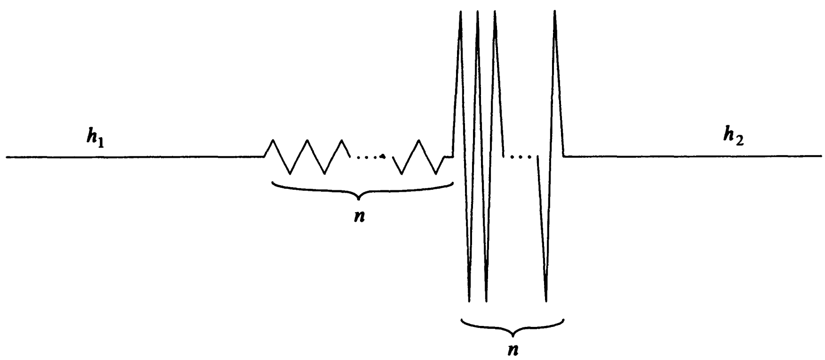



FIG. 13

The path has $n$ "small" sawteeth followed by $n$ "large" sawteeth. The small (resp. large) sawteeth have amplitude $1/n$ (resp. 1) and wavelength $1/n$ (resp. $1/n^2$). Clearly any contact motion of a unit-width ring from one handle $h_1$ to the other handle $h_2$ must make $\Omega(n^2)$ "basic moves" where we may define a basic move to be a motion that is monotone in both $x$ and $y$. One can easily turn this example into a polygon with handles with the property that passing a minimum width ring from one handle to the other requires $\Omega(n^2)$ contact moves. Let us note that this lower bound is invalid for the example if we allow noncontact motion or if we allow an elastic ring; it is then unknown if $\Omega(n^2)$ basic moves can be forced.

Now we address the problem of computing $w_R(P)$. Regard $P$ as the union of two polygonal paths $\gamma_1, \gamma_2$ (the "upper" and "lower" paths). The initial and final portions of the two paths coincide to form the handles. Let the vertex set and edge set of $\gamma_i$ $(i = 1, 2)$ be $V_i$ and $E_i$, respectively. A *position* $p = (x_1, x_2)$ is a pair with $x_1 \in \gamma_1$, $x_2 \in \gamma_2$. The (*combinatorial*) *type* of $p$ is a pair

$$T(p) = (u_1, u_2) \in (V_1 \cup E_1) \times (V_2 \cup E_2),$$

where $x_1 \in u_1$, $x_2 \in u_2$. We regard an edge $e_i \in E_i$ as an open segment, so the combinatorial type of a position is unambiguous. Furthermore, if $h_1, h_2$ are the two handles of $P$, we assume they are in both $E_1$ and $E_2$.

Let $\pi = (\pi_1, \pi_2)$ $(\pi_i\colon [0,1] \to \gamma_i)$ be a contact motion of an elastic ring. The width of $\pi$ is $\sup_t |\pi_1(t) - \pi_2(t)|$. From $\pi$, we get a collection of combinatorial types

$\{T(\pi(t)) \mid t \in [0,1]\}$. We may assume that $\pi$ is well-behaved so that there are a finite number of changes of combinatorial type as $t$ varies. (Otherwise, by standard arguments, we can replace $\pi$ by a well-behaved $\pi'$ with width $(\pi) \geqslant$ width $(\pi')$ and this $\pi'$ suffices for our purposes.) Thus we derive from $\pi$ a finite sequence of combinatorial types

$$T_1, T_2, \ldots, T_k$$

such that the unit interval $[0,1]$ is divided into $k$ time intervals (open, closed or half open)

$$I_1, I_2, \ldots, I_k$$

such that for each $t \in I_i$, $\pi(t)$ is of type $T_i$.

For elements $u, u' \in V_i \cup E_i$, we say $u, u'$ are *adjacent* if either $u = u'$, or $u$ and $u'$ are incident to each other (so that one is a vertex and the other an edge). For combinatorial types $(u, v), (u', v') \in (V_1 \cup E_1) \times (V_2 \cup E_2)$, we say $(u, v)$ and $(u', v')$ are *adjacent* if both $u, u'$ are adjacent and $v, v'$ are adjacent.

For each combinatorial type $(u, v)$, choose a *canonical position* $C(u, v)$ to be any position $(x, y)$ where $x \in \bar{u}$, $y \in \bar{v}$ such that $|x - y|$ is minimized. Here, $\bar{u}$ is the topological closure of $u$, so an edge $u$ becomes a closed segment $\bar{u}$. It is not hard to see that $C(u, v)$ is unique unless $u, v$ are parallel edges, in which case we choose $(x, y)$ so that at least one of $x$ or $y$ is a vertex. Also, we allow $x = y$, a possibility which arises if $u, v$ are adjacent edges.

It is then easy to see:

LEMMA 3. *From any position $(x, y) \in (\gamma_1, \gamma_2)$, there is a motion $\pi_{x,y}$ from $(x, y)$ to $C(T(x, y))$ such that the type of $\pi_{x,y}(t)$ for $t \in (0,1)$ is $T(x, y)$ and the width of $\pi_{x,y}$ is attained at $\pi_{x,y}(0) = (x, y)$.*

Using this property, we may now modify $\pi$ to $\pi'$ so that for each time interval $I_i$, there is a moment $s_i \in I_i$ when $\pi'(s_i) = C(T_i)$. Note that $T(\pi'(s_i))$ need not be equal to $T_i$, but for all $t \in I_i - \{s_i\}$, $\pi'(t)$ has type $T_i$ as before. Furthermore width$(\pi')$ = width$(\pi)$.

Therefore we see that width$(\pi')$ is attained in the portion of $\pi'$ between two canonical positions $\pi'(s_i)$ and $\pi'(s_{i+1})$. By a further modification, we can convert $\pi'$ to $\pi''$ so that width$(\pi'')$ is attained precisely at some canonical position. The basis of this modification is the next lemma.

LEMMA 4. *For any pair of adjacent canonical positions $(x, y)$ to $(x', y')$, there is a "canonical motion" $\pi_{x,y,x',y'}$ from $(x, y)$ to $(x', y')$ whose width is attained as $|x - y|$ or $|x' - y'|$.*

*Proof.* It is easy to see that there is a pair of edges $e$ and $f$ such that $x, x' \in \bar{e}$, $y, y' \in \bar{f}$. Then the motion

$$\pi_{x,y,x',y'}(t) = \big(tx' + (1-t)x, ty' + (1-t)y\big)$$

has the desired property. □

Repeated application of this lemma, by replacing the portion of $\pi'$ between consecutive canonical positions $\pi'(s_i)$ and $\pi'(s_{i+1})$, gives us our final motion $\pi''$.

Furthermore, we assume that $\pi''(0) = (x_0, y_0)$ with $x_0 = y_0$ in $h_1$, and likewise $\pi''(1) = (x_1, y_1)$ with $x_1 = y_1$ in $h_2$. We say that such a $\pi''$ is in *canonical form*.

COROLLARY. *For any polygon $P$ with handles, $w_R(P)$ is equal to the distance between some pair $u, v$ with $u \in V_1 \cup E_1$ and $v \in V_2 \cup E_2$, where at least one of $u$ or $v$ is a vertex.*

We need one more concept before we present the algorithm. Let us define the *initial* and *final* positions. The polygon $P$ has handles $h_1$ and $h_2$. Let $v_i$ be the vertex where $h_i$ attaches to $P$. Then the canonical positions $(v_1, v_1)$ and $(v_2, v_2)$ are called the initial and final positions respectively.

To derive an algorithm, we proceed as follows: we sort the set of all values $|x - y|$ where $(x, y)$ range over all canonical positions. Let the sorted values be

$$0 = r_0 \leqslant r_1 \leqslant \cdots \leqslant r_k,$$

where there are $k$ canonical positions and we break ties indiscriminately. Let $(x_i, y_i)$ be the canonical position associated with $r_i = |x_i - y_i|$. We partition these canonical positions into a collection of disjoint classes. These classes represent an equivalence relation (called *mutual accessibility*) among the canonical positions. Initially, each canonical position is put into its own singleton set, so that they are all inaccessible from each other.

Now we process these canonical positions in stages, where in the $i$th stage $(i = 0, 1, \dots)$, we process $(x_i, y_i)$: when processing $(x_i, y_i)$, we check each canonical position $(x_j, y_j)$ adjacent to $(x_i, y_i)$ (there are at most 6 such). If $j < i$ (so that $(x_j, y_j)$ has already been processed) then we make $(x_i, y_i)$ and $(x_j, y_j)$ mutually accessible. In other words, if they are not yet mutually accessible, then we merge the two classes containing $(x_i, y_i)$ and $(x_j, y_j)$, respectively. We say stage $i$ is *terminal* when the initial position and the final position are mutually accessible. This termination condition is checked just before we start each new stage. The terminal value of $r_i = |x_i - y_i|$ gives us the minimum width.

The justification of this algorithm comes from the following observation:

After we process a canonical position $(x_i, y_i)$, two canonical positions with indices $j_1, j_2 \leqslant i$ are mutually accessible using motions of width at most $r_i$ if and only if they belong to the same class.

Let us briefly note the time complexity of this algorithm: The sorting of the $r_i$'s takes time $O(n^2 \log n)$ since there are $O(n^2)$ canonical positions. To process each $(x_i, y_i)$, we need to maintain a collection of disjoint sets, quickly form the union of two disjoint sets, and determine for any two elements whether they belong to the same set. All of this can be accomplished very efficiently using the well-known *union-find* data structure [**1**]. In particular, any sequence of $t$ such operations takes time $O(t\alpha(m))$ where $m$ is the total number of elements in the sets and $\alpha(m)$ is a very slowly growing function (the so-called inverse Ackermann function). Since $m = O(n^2)$ and $t = O(n^2)$ in our case, the overall time complexity is $O(n^2 \log n)$.

REFERENCES

1. A. V. Aho, J. E. Hopcroft, and J. D. Ullman, The Design and Analysis of Computer Algorithms, Addison-Wesley, Reading, MA, 1974.
2. Fan Chung and Paul Seymour, personal communication.

3. David Eppstein, personal communication.
4. John Kutcher and Joseph O'Rourke, personal communication.
5. J. T. Schwartz and M. Sharir, Efficient motion planning in environments of bounded local complexity, Tech. Rep. No. 164, NYU-Courant Inst.
6. G. Strang, The width of a chair, *Amer. Math. Monthly*, 89 (1982) 529–535.
7. Chee K. Yap, Algorithmic motion planning, in Advances in robotics: Volume 1 (J. T. Schwartz and Chee K. Yap, eds.), Lawrence Erlbaum Assoc's, Hillsdale, NJ, 1987.
8. ______, How to move a chair through a door, *IEEE Trans. Automat. Control*, to appear.
9. ______, How to move a chair through a door, *IEEE J. Robotics and Automation*, RA-3 (1987) 172–181.

---

# LETTERS TO THE EDITOR

Editor,

D. H. Lehmer's article "A New Approach to the Bernoulli Polynomials" [*Monthly*, Dec. 88] makes very pleasant reading but there is a small gap in the presentation. To define the Bernoulli polynomials via Lemma 1 requires that the $b_r$ as given by Equation (11) should be *independent of* $m$. This has not been shown. In fact, Lemmas 1 and 2 show that, *for each fixed* $m$, there exists a unique monic polynomial of degree $n$, say $B_n^{(m)}(x)$, satisfying (6). Since the Fourier series $\phi_n(x)$ of Theorem 4 is *independent of* $m$ and satisfies (6) *for all* $m$, the superscript in $B_n^{(m)}(x)$ can be dropped.

Sincerely,

DAVID CALLAN
University of Bridgeport
Bridgeport, CT 06601

Editor,

While I do wish to thank J. M. Patin for sharing his short proof of Stirling's formula [3] with readers of the MONTHLY, it should be noted that his proof is essentially the same as the one which appears as "Problem #95" in Donald J. Newman's *A Problem Seminar* [**1**; p. 20 (problem), p. 38 (hint), p. 104 (solution)], but in greater detail. In his review of Newman's book, Ivan Niven [**2**] states that this "challenging book for problem buffs" has only one "really misplaced problem," namely this one, requesting "a proof of one form of Stirling's formula, which is surely a textbook matter."

REFERENCES

1. D. J. Newman, A Problem Seminar, Springer-Verlag, New York, 1982
2. I. Niven, A review of *A Problem Seminar*, this MONTHLY, 92 (1985) 437–439.
3. J. M. Patin, A very short proof of Stirling's formula, this MONTHLY, 96 (1989) 41–42.

Yours sincerely,

ROGER B. NELSEN
Lewis & Clark College,
Portland, OR 97219

Editor,

With a great pleasure I read the recent note in The Editor's Corner on a new Mersenne Conjecture (*Amer. Math. Monthly*, 96 (1989) 125–128).

It may be of interest to note that complementary and succinct results hold for Mersenne numbers which are products of two distinct primes. Thus, for a simple example, put $M_n = s^n - 1$. Then $M_n$ is the product of two distinct primes only if $n$ is either a prime $p$ or the square of a prime $q$, in which case precisely one prime factor of $M_n$ is Mersenne, *viz.*, $M_q$. (Note that if $n > 4$ is even then $M_n$ is neither the product of two distinct primes nor prime.)

Analogous results on products of two distinct primes hold for Fibonacci numbers and so-called Repunits, $R_n = (10^n - 1)/9$. Indeed, I suspect that strong primality conjectures of Mersenne type exist for Fibonacci numbers and Repunits, too.

ALBERT A. MULLIN
Huntsville, AL 35806

Editor,

I believe that the following is a simpler argument than the one in the article by Lewin and Lewin (MONTHLY, Dec., 1988, p. 942). The result in question is that if $(a_n)$ is a decreasing sequence of positive numbers, then $\lim a_n = 0$ iff $\Sigma b_n$ diverges, where $b_n = 1 - (a_{n+1}/a_n)$.

*Proof.* First note that, for $M > N$,

$$0 \leqslant \frac{a_N - a_{M+1}}{a_N} = \frac{1}{a_N}\sum_{n=N}^{M}(a_n - a_{n+1}) \leqslant \sum_N^M b_n \leqslant \frac{1}{a_M}\sum_N^M (a_n - a_{n+1})$$
$$= \frac{a_N - a_{M+1}}{a_M}.$$

Now, if $\Sigma b_n$ converges and $a_n \to 0$, then letting $M \to \infty$ gives $1 \leqslant \Sigma_N^\infty b_n$, which is impossible. Conversely, if $a_n \to \alpha > 0$ then let $N = 1$ and let $M \to \infty$. The inequality above yields $\Sigma_1^\infty b_n \leqslant (a_1 - \alpha)/\alpha$, and so the series converges.

ROGER PINKHAM
Stevens Institute of Technology
Hoboken, NJ 07030

# NOTES

EDITED BY DAVID J. HALLENBECK, DENNIS DETURCK, AND ANITA E. SOLOW

## Monotone Multiplicative Functions

JOEL M. COHEN[1]
*University of Maryland, College Park, MD 20742*

The purpose of this note is to present a simple and intuitive proof of the following classical result [2] about multiplicative functions defined on the natural numbers $\mathbb{N}$. Recall first that a function $f\colon \mathbb{N} \to \mathbb{R}^+$ is **multiplicative** if $f(mn) = f(m)f(n)$ whenever $m$ and $n$ are relatively prime, i.e., if the greatest common divisor, $(m, n)$, is 1.

THEOREM. *If $f\colon \mathbb{N} \to \mathbb{R}^+$ is a monotone multiplicative function, then there exists some $r$ such that $f(n) = n^r$ for all $n$.*

Reference [3] contains a discussion of the problem and of some earlier proofs. Our proof is more conceptual than those listed in the bibliography and is based on one simple observation: a multiplicative function $f\colon \mathbb{N} \to \mathbb{R}^+$ can be extended uniquely to a "multiplicative function" on $\mathbb{Q}^+$. If $f$ is increasing on $\mathbb{N}$, then it is also increasing on $\mathbb{Q}^+$ and can be extended to an increasing function on $\mathbb{R}^+$, which is almost totally multiplicative. It is then easy to prove that $f$ is continuous and totally multiplicative, and it is well known that a continuous endomorphism of $\mathbb{R}^+$ is necessarily raising to a power. For the proof, we need the following well-known fact from real analysis: An increasing function from the reals to themselves has a countable set of discontinuities. Except for this, the paper is self-contained.

First, whenever we write a rational number $p/q \in \mathbb{Q}^+$, it will be assumed that $p$ and $q$ are positive integers such that $(p, q) = 1$. Two rational numbers $\alpha = p/q$ and $\beta = r/s$ are **relatively prime**, written $(\alpha, \beta) = 1$, if $p$, $q$, $r$, and $s$ are relatively prime in pairs. A function $f\colon \mathbb{Q}^+ \to \mathbb{R}^+$ is called **multiplicative** if $f(\alpha\beta) = f(\alpha)f(\beta)$ whenever $(\alpha, \beta) = 1$.

LEMMA. *Every given positive rational number $p/q$ (and hence every given positive real number) can be approximated arbitrarily closely from above and from below by rational numbers that are relatively prime to any given finite set of positive rational numbers.*

*Proof.* Let $S$ be the given finite set, and let $m$ be the product of $pq$ and all the numerators and denominators in $S$. Let $n$ be an arbitrary positive integer. Then the two rational numbers $r_\varepsilon = (pnm + \varepsilon)/(qnm - \varepsilon)$, where $\varepsilon = \pm 1$, are relatively prime to $p/q$ and to each rational number in $S$, and they satisfy the inequalities $r_{-1} < p/q < r_{+1}$. Moreover each differs from $p/q$ by less than $(p + q)/(qnm - 1)$, which can be made arbitrarily small by choosing $n$ large enough.

Let $f\colon \mathbb{N} \to \mathbb{R}^+$ be an increasing multiplicative function. For $\alpha = p/q \in \mathbb{Q}^+$, define $f(\alpha) = f(p)/f(q)$. It is immediate that $f\colon \mathbb{Q}^+ \to \mathbb{R}^+$ is multiplicative. We now show that it is also increasing: If $\alpha = p/q < \beta = r/s$ and if $(\alpha, \beta) = 1$, then $sp < rq$, so that $f(s)f(p) = f(sp) \le f(rq) = f(r)f(q)$, whence $f(\alpha) \le f(\beta)$. If $\alpha$

[1] Partially supported by the National Science Foundation.

and $\beta$ are not relatively prime, by the lemma we can find $\gamma$ such that $\alpha < \gamma < \beta$, and $(\alpha, \gamma) = (\gamma, \beta) = 1$. Thus $f(\alpha) \leq f(\gamma) \leq f(\beta)$. So $f$ is increasing and multiplicative. We can extend it to $\mathbb{R}^+$ as follows: for $\eta$ irrational, let $f(\eta) = \lim_{\alpha \to \eta^-} f(\alpha)$, the limit taken from $\alpha \in \mathbb{Q}^+$. Note that $f$ is increasing and is continuous from the left at irrationals.

Let $\eta, \theta \in \mathbb{R}^+$ with $\theta$ irrational. Using the lemma, we can find $\alpha$ and $\beta$ relatively prime rational numbers less than but arbitrarily close to $\eta$ and $\theta$ respectively, except that if $\eta$ is rational we take $\alpha = \eta$. Then $f(\alpha)f(\beta) = f(\alpha\beta)$. Passing to limits we get $f(\eta)f(\theta) = f(\eta\theta)$. Thus $f$ is multiplicative except possibly on the product of rational numbers not relatively prime.

Now $f$ is increasing on $\mathbb{R}^+$, so it has a countable set of discontinuities. In particular, if $\delta$ is any real number, then we can find an irrational number $\xi$ such that $f$ is continuous at $\delta\xi$. Since $\xi$ is irrational, $f(\alpha) = f(\alpha\xi)/f(\xi)$ for any real number $\alpha$. Thus $\lim_{\alpha \to \delta} f(\alpha) = \lim_{\alpha \to \delta} f(\alpha\xi)/f(\xi) = f(\delta\xi)/f(\xi) = f(\delta)$. Thus $f$ is continuous everywhere. If $\alpha$ and $\beta$ are rational, then approach $\beta$ by irrational numbers $\theta$. Since $f$ is continuous, $f(\alpha\beta) = \lim_{\theta \to \beta} f(\alpha\theta) = \lim_{\theta \to \beta} f(\alpha)f(\theta) = f(\alpha)f(\beta)$, so $f$ is totally multiplicative. Thus $\log \circ f \circ \exp$: $\mathbb{R} \to \mathbb{R}$ is a continuous homomorphism and so by elementary linear algebra is multiplication by some constant $r$. So $f(x) = x^r$ and the theorem is proved for the case of $f$ increasing. For $f$ decreasing $1/f$ is increasing and we obtain the same result.

The author is grateful to the referee for improving the readability of this note.

REFERENCES

1. A. S. Besicovitch, On additive functions of a positive integer, Studies in Math. Anal. and Related Topics, Stanford University Press, Stanford, CA, 1962.
2. P. Erdös, On the distribution function of additive functions, *Ann. of Math.*, (2) 47(1946) 1–20.
3. E. Howe, A new proof of Erdös's Theorem on monotone multiplicative functions, *Amer. Math. Monthly*, 93(1986) 593–595.
4. L. Moser and J. Lambek, On monotone multiplicative functions, *Proc. Amer. Math. Soc.*, 4(1953) 544–545.
5. I. J. Schoenberg, On two theorems of P. Erdös and A. Rényi, *Ill. J. Math.*, 6(1962) 53–58.

## Powers of a Prime Dividing Binomial Coefficients

W. J. WONG
*Department of Mathematics, University of Notre Dame, Notre Dame, IN 46556*

The binomial coefficients

$$C(n, k) = \frac{n!}{k!(n-k)!}$$

form an enduring source of interesting problems and results. In a recent MONTHLY article **[1]**, Goetgheluck explained how patterns calculated with the aid of a microcomputer led him to rediscover a result of Kummer [3] giving the exact power of a prime $p$ dividing $C(n, k)$, enabling him to give a fast way of computing $C(n, k)$ for large values of $n$. In this note, I give a result about the set of powers of $p$ exactly dividing the $C(n, k)$, for $0 \leqslant k \leqslant n$, suggested by thinking about a

situation in the theory of representations of algebraic groups (see remarks near the end of the paper). Computers also enter the story, but as an agent of motivation rather than discovery.

First we introduce some notation and terminology. If $a$, $b$ are integers, $a \leqslant b$, we write $[a, b]$ for the *interval* of integers

$$[a, b] = \{x \in Z | a \leqslant x \leqslant b\}.$$

If $I = [a, b]$, we denote the length $b - a$ of the interval by $|I|$. Let $p$ be a prime number, and $v_p$ the corresponding valuation of the ring of integers, that is, if $m$ is a nonzero integer, then $v_p m = r$, where $p^r$ is the largest power of $p$ dividing $m$. Given a positive integer $n$, let

$$E(p, n) = \{v_p C(n, k) | k \in [0, n]\}.$$

The question suggested by the algebraic group problem was: What is the nature of this set of integers? The simplest possible answer would be that it is an interval, and I verified this for small values of $n$ by writing out Pascal's triangle. However, this was not much evidence, since the bottom of the blackboard was reached at $n = 12$, by which time no $v_p C(n, k)$ had exceeded 3.

At this point, computers came into the picture. I mentioned the question to a number of colleagues, and three of them went away and, independently, wrote programs to determine whether the sets $E(p, n)$ were intervals, for values of $n$ up to quite a high number. (One of them, Steve Doty, left his Commodore 64 running all night, reaching a value of $n$ above 500, without finding any case in which $E(p, n)$ was not an interval.) It was clearly time to state a theorem and find a proof.

THEOREM. *If $p$ is a prime number and $n$ a positive integer, then the set*

$$E(p, n) = \{v_p C(n, k) | k \in [0, n]\}$$

*is an interval of integers.*

*Proof.* To make the proof self-contained, we demonstrate some more or less well-known formulas. In particular, we give an argument for the result of Kummer and Goetgheluck which will enable us to make a construction necessary for the proof.

If $m$ is any nonnegative integer, we let $m(i)$ denote the $i$th digit in the expansion of $m$ in base $p$:

$$m = \sum_{i \geqslant 0} m(i) p^i, \qquad 0 \leqslant m(i) \leqslant p - 1.$$

Since the number of integers $x$ in $[1, m]$ having $v_p x \geqslant j$ (for $j \geqslant 1$) is clearly $[m/p^j]$, where $[\ ]$ denotes the greatest integer function, we have Legendre's formula

$$v_p(m!) = \sum_{j \geqslant 1} [m/p^j].$$

From the equations

$$[m/p^j] = \sum_{i \geqslant j} m(i) p^{i-j},$$

$$\sum_{j=1}^{i} p^{i-j} = (p^i - 1)/(p - 1),$$

we find that

$$v_p(m!) = \left(m - \sum_{i\geqslant 0} m(i)\right)/(p-1).$$

If $0 \leqslant k \leqslant n$, $h = n - k$, we obtain a formula of Kummer

$$v_pC(n,k) = \sum_{i\geqslant 0} (k(i) + h(i) - n(i))/(p-1). \tag{A}$$

If $k(i) \leqslant n(i)$, for all $i$, then there are no "borrows" when $k$ is subtracted from $n$ in base $p$, so that $h(i) = n(i) - k(i)$, and $v_pC(n,k) = 0$. Otherwise, borrows arise in the following way. Let $b$ be the largest integer for which $n(b) > 0$, let $s$ be the smallest integer for which $k(s) > n(s)$, and let $t$ be the smallest integer greater than $s$ for which $k(t) < n(t)$. (Since $k \leqslant n$, $t$ exists.) We obtain a subinterval $I = [s, t]$ of $[0, b]$, such that

$$k(s) > n(s), \qquad k(x) \geqslant n(x) \quad \text{when } s < x < t, \quad \text{and } k(t) < n(t).$$

Repeating this procedure on the interval $[t+1, b]$, we obtain a collection $\Gamma$ of disjoint subintervals with the above property, such that $k(i) \leqslant n(i)$ if $i$ does not lie in one of the subintervals. From the equation

$$\sum_{i\in I} n(i)p^i - \sum_{i\in I} k(i)p^i = (p + n(s) - k(s))p^s + \sum_{s<x<t} (p - 1 + n(x) - k(x))p^x + (-1 + n(t) - k(t))p^t,$$

we see that there are $t - s$ borrows in places corresponding to $I$, and also that

$$\sum_{i\in I} (k(i) + h(i) - n(i))/(p-1) = t - s.$$

If $i$ does not lie in a subinterval in $\Gamma$, there is no borrow in place $i$. It now follows from (A) that

$$v_pC(n,k) = \text{number of borrows in subtracting } k \text{ from } n \text{ in base } p, \tag{B}$$

the result of Kummer rediscovered by Goetgheluck.

If $n(i) = p - 1$, for all $i$ less than $b$, then there can be no borrows, so that $v_pC(n,k) = 0$, for all $k$ in $[0, n]$. Suppose this is not the case, and let $a$ be the smallest integer such that $n(a) < p - 1$. Then the largest possible number of borrows in subtracting $k$ from $n$ is $b - a$, since there can be no borrows in places less than $a$. Now let

$$S = \{a\} \cup \{i \mid a < i \leqslant b,\ n(i) = 0\},$$
$$T = \{i \mid a < i \leqslant b,\ n(i) > 0\}.$$

Then $a \in S$, $b \in T$, and $[a, b]$ is the disjoint union of $S$ and $T$.

Now suppose that $\Gamma$ is a collection of disjoint subintervals of $[a, b]$, such that each member $I$ of $\Gamma$ has the form $I = [s, t]$, where $s \in S$, $t \in T$. For such an $I$ form the number

$$(n(s) + 1)p^s + \sum_{s<x<t} n(x)p^x + (n(t) - 1)p^t,$$

and let $k$ be the sum of all these numbers, as $I$ ranges over $\Gamma$. From the argument used to show (B), we see that

$$v_p C(n, k) = \sum_{I \in \Gamma} |I|.$$

To complete the proof of the theorem, it suffices to prove the following:

PROPOSITION. *Suppose an interval of integers* $[a, b]$ *is a disjoint union of subsets* $S$ *and* $T$, *where* $a \in S$, $b \in T$. *If* $1 \leqslant r \leqslant b - a$, *then there exists a collection* $\Gamma$ *of disjoint subintervals of* $[a, b]$, *with the property that each member* $I$ *of* $\Gamma$ *has the form* $[s, t]$, *where* $s \in S$, $t \in T$, *such that*

$$\sum_{I \in \Gamma} |I| = r.$$

*Proof.* Clearly $[a, b]$ splits into intervals

$$S_i = [a_i, c_i] \subset S, \quad T_i = [c_i + 1, b_i] \subset T \qquad (1 \leqslant i \leqslant m),$$

such that $a_1 = a$, $b_m = b$, $a_{i+1} = b_i + 1$ $(1 \leqslant i < m)$.

First, suppose $r > b - a - m$. If $j = r - b + a + m$, then $1 \leqslant j \leqslant m$. Take $\Gamma$ to be the collection of intervals $[a, b_j], [a_{j+1}, b_{j+1}], \ldots, [a_m, b_m]$.

Now suppose $r \leqslant b - a - m$. Since

$$\sum_{i=1}^{m} (b_i - a_i) = b - a - m + 1,$$

we can find integers $r_1, \ldots, r_m$, such that

$$\sum_{i=1}^{m} r_i = r, \qquad 0 \leqslant r_i \leqslant b_i - a_i.$$

For $r_i > 0$, we can clearly choose an interval $[s_i, t_i]$ with $s_i \in S_i$, $t_i \in T_i$, $t_i - s_i = r_i$. Take $\Gamma$ to be the set of intervals obtained in this way. This completes the proof of the proposition, and so also the proof of the theorem.

The problem dealt with here originated in a question in the theory of representations of semisimple algebraic groups $G$ over a field $K$ of characteristic $p$. Reduction modulo $p$ of irreducible modules in the characteristic 0 situation gives a class of modules for $G$, called Weyl modules, because their dimensions are given by Weyl's dimension formula. Using a certain "contravariant" bilinear form, Jantzen defined an important filtration for each Weyl module [2].

In the simplest case $G = \mathrm{SL}_2(K)$, the Weyl modules can be found explicitly. Each element of $G$ gives an algebra automorphism of the polynomial algebra $K[X, Y]$ by means of a linear transformation of the variables $X$, $Y$. In this way, $K[X, Y]$ becomes a module for $G$ over $K$. If $n$ is any nonnegative integer, the homogeneous polynomials of degree $n$ form a submodule $V^{(n)}$, and this is the dual of a Weyl module for $G$. A basis for $V^{(n)}$ is given by the monomials $X^k Y^{n-k}$, where $k \in [0, n]$. The (dual) Jantzen filtration for $V^{(n)}$ is then given by

$$\{0\} = V_0^{(n)} \subset V_1^{(n)} \subset V_2^{(n)} \subset \cdots \subset V^{(n)},$$

where $V_j^{(n)}$ is the subspace spanned by all the $X^k Y^{n-k}$ for which $v_p C(n, k) < j$. A natural question is whether the sequence of submodules is *strictly* increasing (until we reach $V^{(n)}$). Our theorem shows that it is.

The sets $E(p, n)$ can be defined even when $p$ is not a prime number. However, my colleague Dennis Snow has used a computer to find examples such as $E(6,115) = \{0,1,2,3,4,6\}$ which are not intervals. What can be said about the nature of the $E(p, n)$ when $p$ is composite? Perhaps the computer will show us what theorem we should try to prove in this case.

REFERENCES

1. P. Goetgheluck, Computing binomial coefficients, *Amer. Math. Monthly*, 94 (1987) 360–365.
2. J. C. Jantzen, Weyl modules for groups of Lie type, *Finite Simple Groups II*, Proc. Durham 1978, London/New York (Academic Press) 1980.
3. E. E. Kummer, Über die Ergänzungssätze zu den allgemeinen Reziprozitätsgesetzen, *J. reine u. angew. Math.*, 44 (1852) 93–146.

# The Polar Decomposition and a Matrix Inequality

DERMING WANG
*Department of Mathematics and Computer Science, California State University, Long Beach, CA 90840*

A simple observation and an interesting computer experiment led Chan and Kwong [1] to a conjecture about a pair of matrix inequalities. Their conjecture was recently solved by Furuta [2], who proved a pair of more general operator inequalities that contain, as a special case, the Chan-Kwong conjecture. In this note I shall present an alternate elementary and self-contained proof of this conjecture, based on the polar decomposition of a matrix. First I will state some relevant definitions and, for motivation, give a background account of the work that led to the conjecture. I also state the Furuta inequalities.

All matrices considered here are assumed to be $n \times n$ complex matrices acting on an $n$-dimensional inner product space with inner product $\langle \cdot, \cdot \rangle$. A matrix $T$ is said to be Hermitian if $T^* := T$, where $T^*$ is the adjoint of $T$. A Hermitian matrix is said to be nonnegative, in notation $T \geqslant 0$, if $\langle Tx, x\rangle \geqslant 0$ for every vector $x$. For Hermitian matrices $S$ and $T$, write $S \geqslant T$ or $T \leqslant S$ if $S - T \geqslant 0$. Every nonnegative matrix $T$ possesses a (unique, nonnegative) square root, and, I shall denote this square root $T^{1/2}$.

The following facts are well known.

(1) If $S \geqslant T$, then $W^*SW \geqslant W^*TW$ for any matrix $W$. This follows from the definition of "$\geqslant$" and the definition of adjoint.

(2) If $S \geqslant T \geqslant 0$, then $S^{1/2} \geqslant T^{1/2}$. Thus the square-root function is order-preserving.

(3) It is not true that $S \geqslant T \geqslant 0$ implies $S^2 \geqslant T^2$. Thus the square function is not order-preserving.

Convenient references for (2) and (3) are [1] and [3]. In particular, [3] contains an elementary and elegant proof of (2) discovered by Kato.

Consider the matrix inequality $A \geqslant B$ with $A$ and $B$ nonnegative. When the square function is applied to both sides of the inequality the order will, in general, be destroyed. The original order is nevertheless restored if the square-root function follows afterward. This led the authors of [1] to investigate whether the original

order is preserved if the operation "multiplying both on the left and on the right by a matrix $C \geqslant 0$" is performed between the square function and the square-root function. The question is, thus, whether the inequality $(CA^2C)^{1/2} \geqslant (CB^2C)^{1/2}$ is valid. Although (1) seems to indicate that this "middle" operation should not have worsened the situation, the authors were able to discover a counterexample with the help of the MAT-LAB package on a VAX computer [**1**]. In this counterexample, $A$, $B$, and $C$ are $2 \times 2$ matrices satisfying $A \geqslant B \geqslant C \geqslant 0$.

This same computer experiment seems to indicate that the inequality persists under the special circumstances $C = A$ or $C = B$. Since the authors were able to verify that this indeed is the case when $A$ and $B$ are $2 \times 2$ matrices, they, therefore, proposed the following conjecture: If $A \geqslant B \geqslant 0$, then $(BA^2B)^{1/2} \geqslant B^2$ and $A^2 \geqslant (AB^2A)^{1/2}$.

For bounded linear operators $A$ and $B$ acting on a Hilbert space, Furuta proved that if $A \geqslant B \geqslant 0$, then $(B^rA^pB^r)^{1/q} \geqslant B^{(p+2r)/q}$ and $A^{(p+2r)/q} \geqslant (A^rB^pA^r)^{1/q}$, where $p, r \geqslant 0$, $q \geqslant 1$ and $(1 + 2r)q \geqslant p + 2r$. With $r = 1$ and $p = q = 2$, these inequalities reduce to those conjectured by Chan and Kwong.

I am ready to present the alternate proof by establishing the following:

THEOREM. *If* $A \geqslant C \geqslant B \geqslant 0$, *then* $(CA^2C)^{1/2} \geqslant (CB^2C)^{1/2}$.

It is readily seen that the statement of the theorem is equivalent to the conjecture. The form of the inequality seems to suggest some connection with the polar decomposition of a matrix. Indeed, the novel idea of the proof is based on the decomposition of the matrices $A^{1/2}C^{1/2}$ and $B^{1/2}C^{1/2}$. Recall that a matrix $U$ is said to be unitary if $U^*U = UU^* = I$, where $I$ is the identity matrix. The polar decomposition states that for any matrix $T$, there is a (not necessarily unique) unitary matrix $U$ such that $T = U(T^*T)^{1/2}$. A proof of this decomposition may be found in [**4**].

*Proof.* Let $H$ be the square root of $(A^{1/2}C^{1/2})^*(A^{1/2}C^{1/2}) = C^{1/2}AC^{1/2}$. Thus there is a unitary matrix $U$ such that $A^{1/2}C^{1/2} = UH$ and $C^{1/2}A^{1/2} = HU^*$. Since $A \geqslant C$, $H^2 = C^{1/2}AC^{1/2} \geqslant C^2$, and hence $H \geqslant C$. Similarly, $A \geqslant C$ implies

$$\begin{aligned} A^2 &\geqslant A^{1/2}CA^{1/2} \\ &= (A^{1/2}C^{1/2})(C^{1/2}A^{1/2}) \\ &= UH^2U^*. \end{aligned}$$

Thus, $A \geqslant UHU^*$. Therefore, $U^*AU \geqslant H \geqslant C$.

Now,

$$\begin{aligned} CA^2C &= C^{1/2}(C^{1/2}A^{1/2})A(A^{1/2}C^{1/2})C^{1/2} \\ &= C^{1/2}H(U^*AU)HC^{1/2} \\ &\geqslant C^{1/2}HCHC^{1/2} \\ &= (C^{1/2}HC^{1/2})^2. \end{aligned}$$

Taking square roots, we obtain

$$(CA^2C)^{1/2} \geqslant C^{1/2}HC^{1/2} \geqslant C^{1/2}CC^{1/2} = C^2.$$

If we retrace the preceding steps replacing each occurrence of $A$ by $B$ and "$\geqslant$" by "$\leqslant$", we obtain $(CB^2C)^{1/2} \leqslant C^2$. The theorem now follows by transitivity.

*Remarks.* (i) As was noted in [**1**], the inequalities conjectured by Chan and Kwong and proved by Furuta may be used as an alternate approach to prove the following [**1**, Theorem 5]: If $B$ and $Q$ are nonnegative matrices with $B$ invertible, then the solution matrix $X$ of the equation $B^2X + XB^2 = BQ + QB$ must be nonnegative.

(ii) The inequalities proved by Furuta are, of course, valid for operators. A slight modification shows the proof presented above is also valid for operators. Indeed, in the operator setting, (1) and (2) still hold but the operator $U$ in the polar decomposition will, in general, be just a partial isometry. However, one may choose a partial isometry $U$ in the decomposition of $A^{1/2}C^{1/2}$ so that $U$ has the same kernel as $H$. In this case, $U^*U$ is the orthogonal projection onto the norm closure of the range of $H$. (See, for example [**5**].) Thus $U^*UH = H = U^*UHU^*U$. Therefore, the square root of $UH^2U^*$ is $UHU^*$ and $A \geqslant UHU^*$ implies $U^*AU \geqslant H$. The remainder of the proof goes through without change.

**Acknowledgment.** The author wishes to express his gratitude to the referee for calling his attention to the Furuta paper.

REFERENCES

1. N. N. Chan and M. K. Kwong, Hermitian matrix inequalities and a conjecture, this MONTHLY, 92(1985) 533–541.
2. T. Furuta, $A \geqslant B \geqslant 0$ assures $(B^rA^pB^r)^{1/q} \geqslant B^{(p+2r)/q}$ for $r \geqslant 0$, $p \geqslant 0$, $q \geqslant 1$ with $(1 + 2r)q \geqslant p + 2r$, *Proc. Amer. Math. Soc.*, 100(1987) 1–3.
3. C. Davis, Notions generalizing convexity for functions defined on spaces of matrices, *Proceedings of Symposia in Pure Mathematics, Amer. Math. Soc.*, 7(1963) 187–201.
4. I. Kaplansky, Linear Algebra and Geometry, Allyn and Bacon, Boston, 1969.
5. R. G. Douglas, Banach Algebra Techniques in Operator Theory, Academic Press, New York, 1972.

# A Natural Interpretation of an Artificial Function

HANSKLAUS RUMMLER
*Institut de Mathématiques de l'Université de Fribourg, CH-1700 Fribourg, Switzerland*

Every student of introductory analysis learns the example of the real function

$$f(t) = \begin{cases} 0 & \text{if } t \text{ is irrational,} \\ \dfrac{1}{|q|} & \text{if } t = \dfrac{p}{q} \text{ rational with } \gcd(p, q) = 1, \end{cases}$$

which is intended to show that the concept of continuity is more subtle than one expects, because $f$ is continuous exactly at the irrational points (see [**1**, p. 86] or [**2**, p. 94]). And for the average student, this is the only property of the function $f$, that is, it is not interesting otherwise because there seems to be no reasonable context in which this function arises naturally. Its strange continuity behavior is only just punishment for having created a function by an artificial {-definition!

But there is an interesting interpretation of this function: We consider only positive arguments, and we interpret $t$ as the slope of a ray in the first quadrant of the $x$-$y$ plane, starting at the origin. Let us intercept the ray at $x = n$, where $n$ is any natural number chosen "at random." Then $f(t)$ is the "probability" that the $y$-coordinate of the interception point is also an integer, because this is the case if and only if $n$ is divisible by $q$, and thus in making a great number of random guesses of the integer $x$-coordinate $n$ we shall have also an integer $y$-coordinate in one of $q$ cases. (Strictly speaking, there is no probability measure on $\mathbb{N}$ that makes the choice of each natural number equally probable! But if you choose your number $n$ among the natural numbers $1, 2, \ldots, N$, randomly with respect to the uniform probability measure, your chance to get one which is divisible by $q$ is $[N/q]/N$, which tends to $1/q$ for $N \to \infty$. So $f(t) = 1/q$ is the asymptotic density that $n$ chosen in $\mathbb{N}$ is divisible by $q$.) Of course, this interpretation of the function $f$ may be reformulated for a torus instead of the $x$-$y$ plane.

A slight modification of the function $f$ doesn't change its continuity behavior but admits another nice interpretation: Place a square billiard table in the first quadrant of the $x$-$y$ plane, with its lower-left corner at the origin, and let a ball start at this corner and proceed along the line with slope $t > 0$. Then

$$g(t) = \begin{cases} 0 & \text{if } t \text{ is irrational,} \\ \dfrac{1}{p + q - 1} & \text{if } t = \dfrac{p}{q} \text{ rational with } \gcd(p, q) = 1, \end{cases}$$

is the "probability" that at the $n$th reflection at the boundary, $n$ chosen "at random," the ball is just in a corner.

Unfortunately, these interpretations of the functions $f$ and $g$ do not yet explain their strange continuity behavior, but they show that these functions are not as artificial as students may think if they know only the formal definition.

REFERENCES

1. W. Rudin, Principles of Mathematical Analysis, second ed., McGraw-Hill, New York, 1964.
2. M. Spivak, Calculus, W. A. Benjamin, New York, 1967.

# Factor Rings of Integers

STEVE JOHNSON
*Mathematics Department, Seattle Pacific University, Seattle, WA 98119*

One of the problems in Fraleigh's text, *A First Course in Abstract Algebra*, asks whether or not $2\mathbb{Z}/8\mathbb{Z}$ and $\mathbb{Z}/4\mathbb{Z}$ are isomorphic as rings. A quick look at the multiplication tables verifies that they are not, and thus we have two nonisomorphic rings of order 4. A little more work shows that $4\mathbb{Z}/16\mathbb{Z}$ is yet another ring of order 4. A natural question presents itself: Are these the only rings of order 4 which arise in this way? Further, what about the general case of $n\mathbb{Z}/nk\mathbb{Z}$? This note answers the question, and in so doing provides a large set of examples which can be used to help illuminate the idea of factor rings.

The main result is the following:

THEOREM. *Let $k$ be an integer greater than or equal to 2. Let $d_1, d_2, \ldots, d_m$ be a complete list of the divisors (not necessarily proper) of $k$. Then the $m$ rings $d_1\mathbb{Z}/d_1k\mathbb{Z}$, $d_2\mathbb{Z}/d_2k\mathbb{Z}, \ldots, d_m\mathbb{Z}/d_mk\mathbb{Z}$ are nonisomorphic and form a complete list of the rings of the form $n\mathbb{Z}/nk\mathbb{Z}$.*

In the above example, $k$ is 4. Since the only divisors of 4 are 1, 2, and 4, the list of rings given is in fact the complete list. The theorem is an immediate corollary of the following two lemmas.

LEMMA 1. *If $d_1$ and $d_2$ are distinct divisors of $k$, then $d_1\mathbb{Z}/d_1k\mathbb{Z}$ and $d_2\mathbb{Z}/d_2k\mathbb{Z}$ are not isomorphic.*

*Proof.* Without loss of generality assume that $d_1 < d_2$, and let $s = k/d_2$. Since $d_1 < d_2$, $d_1s < d_2s = k$, hence $d_1^2s < d_1k$, and so $d_1^2s$ is not congruent to 0 modulo $d_1k$.

Notice that for a function $f\colon d_1\mathbb{Z}/d_1k\mathbb{Z} \to d_2\mathbb{Z}/d_2k\mathbb{Z}$ to be an isomorphism a necessary (but not sufficient!) condition is that there exists an $r$ such that $f(d_1i + d_1k\mathbb{Z}) = d_2ri + d_2k\mathbb{Z}$. For such an $f$ we have $f(d_1 + d_1k\mathbb{Z})f(d_1s + d_1k\mathbb{Z}) = (d_2r + d_2k\mathbb{Z})(d_2rs + d_2k\mathbb{Z}) = r^2d_2k + d_2k\mathbb{Z} = 0 + d_2k\mathbb{Z}$. On the other hand $(d_1 + d_1k\mathbb{Z})(d_1s + d_1k\mathbb{Z}) = d_1^2s + d_1k\mathbb{Z} \neq 0 + d_1k\mathbb{Z}$. Thus such an $f$ cannot preserve multiplication and so the rings are not isomorphic.

Following common practice, the notation $(a, b)$ will be used to denote the greatest common divisor of $a$ and $b$.

LEMMA 2. *Let $d = (n, k)$. Then $d\mathbb{Z}/dk\mathbb{Z}$ and $n\mathbb{Z}/nk\mathbb{Z}$ are isomorphic as rings.*

*Proof.* We want a function $f\colon d\mathbb{Z}/dk\mathbb{Z} \to n\mathbb{Z}/nk\mathbb{Z}$ that is a ring isomorphism. First notice that $f$ will be a group homomorphism iff it is of the form $f(di + dk\mathbb{Z}) = nri + nk\mathbb{Z}$ where $(r, k) = 1$. If $f$ is to also be a ring isomorphism, we must in addition have $f((di + dk\mathbb{Z})(dj + dk\mathbb{Z})) = f(di + dk\mathbb{Z})f(dj + dk\mathbb{Z})$. Using the

above form for $f$, this becomes $nrdij + nk\mathbb{Z} = n^2r^2ij + nk\mathbb{Z}$, or equivalently $n^2r^2ij - nrdij \in nk\mathbb{Z}$. Notice that this will hold for all $i$ and $j$ iff it holds for $i = j = 1$; thus we can reduce our condition to the requirement that $r(nr - d) \in k\mathbb{Z}$. Since $d = (n, k)$, there must exist $s$ and $t$ such that $(s, t) = 1$, $n = ds$ and $k = dt$. Substituting $ds$ for $n$ in the condition gives $r(dsr - d) \in dt\mathbb{Z}$, which reduces to $r(sr - 1) \in t\mathbb{Z}$.

To summarize, we know we have an isomorphism if we can find an $r$ such that $(r, k) = 1$ and $r(sr - 1) \in t\mathbb{Z}$. But now $(s, t) = 1$; thus $s$ is a unit modulo $t$, and so there is an $r'$ such that $sr' - 1 \in t\mathbb{Z}$. Consider the set of integers of the form $r' + it$. Since $r'$ is a unit modulo $t$, $(r', t) = 1$. By a theorem of Dirichlet, this set must contain infinitely many primes. Let $r$ be a prime in this set which is larger than $k$. Then $(r, k) = 1$ is clear, and since $r$ and $r'$ are congruent modulo $t$, $sr - 1$, and so $r(sr - 1)$, is in $t\mathbb{Z}$. With this $r$, the map $f(di + dk\mathbb{Z}) = nri + nk\mathbb{Z}$ gives the required isomorphism.

## A Simple Estimate of the Error in Linear Approximation

ROBERT M. GETHNER
*Department of Mathematics and Astronomy, Franklin and Marshall College, Lancaster, PA 17604*

One of the most important topics in calculus, in view of the increasing use of numerical methods, is Taylor's formula with remainder. Unfortunately, this is also one of the most difficult topics to teach: the formula is complicated to apply and its proofs are difficult to motivate. I suggest that we prepare our students for these difficulties by first teaching remainder estimation in the context of *linear* approximation. The following argument, though it yields a slightly weaker estimate than usual, is easy to motivate and demonstrates the power of the mean value theorem. It can be presented quite early, immediately after the mean value theorem and the definition of the second derivative. It also provides the starting point for a proof of a general Taylor remainder estimate.

THEOREM. *If $f''$ exists and is bounded on the closed interval $I$ with endpoints $a$ and $x$, then*

$$f(x) = f(a) + f'(a)(x - a) + R,$$

*where*

$$|R| < M(x - a)^2,$$

*and $M$ is the maximum of $|f''|$ on $I$.*

The usual estimate is half that given by the Theorem.

*Proof.* By the mean value theorem,

$$f(x) = f(a) + f'(c)(x - a),$$

where $c$ is between $a$ and $x$. But by the mean value theorem once again,

$$f'(c) = f'(a) + f''(c^*)(c - a),$$

where $c^*$ is between $a$ and $c$. Hence

$$f(x) = f(a) + f'(a)(x - a) + f''(c^*)(c - a)(x - a).$$

Since $|c - a| < |x - a|$, the theorem is proved.

Linear approximation provides motivation for the general Taylor formula as follows. In the linear setting, $f$ is approximated by a function with constant derivative. This suggests that we could get a better estimate on $f$ by using a linear approximation to its derivative:

$$f'(t) \approx f'(a) + f''(a)(t - a) \quad (\text{all } t \text{ between } a \text{ and } x).$$

Integrating both sides with respect to $t$ from $a$ to $x$ yields

$$f(x) \approx f(a) + f'(a)(x - a) + (1/2)f''(a)(x - a)^2,$$

and the right-hand side is the second degree Taylor polynomial. When this process is repeated (beginning with a quadratic approximation to the derivative), the higher-degree polynomials emerge. The same argument, but beginning with the Theorem and applying the triangle inequality to the integrated remainder, yields corresponding error estimates. (This derivation is essentially that of Nicholas [**2**, **3**]; a recent paper of Brauer [**1**] suggested the idea of deriving an upper bound on $|R|$ without getting an explicit formula for $R$.)

*Acknowledgement.* This paper has benefited greatly from the valuable comments of Prof. G. Rosenstein and the referee.

REFERENCES

1. F. Brauer, A simplification of Taylor's theorem, this MONTHLY, 94 (1987) 453–455.
2. C. P. Nicholas, Taylor's theorem in a first course, this MONTHLY, 58 (1951) 559–562.
3. ______, More on Taylor's theorem in a first course, this MONTHLY, 60 (1953) 329–331; reprinted in the MAA collection, Selected Papers on Calculus, ed. Tom M. Apostol, et al., 1969.

# PROBLEMS AND SOLUTIONS

EDITED BY PAUL T. BATEMAN, HAROLD G. DIAMOND, KENNETH B. STOLARSKY (ADVANCED PROBLEMS), AND DOUGLAS B. WEST (ELEMENTARY PROBLEMS)



*For instructions about submitting **proposed** problems for publication in this department see the inside front cover. Please include solutions, relevant references, etc. Three copies are requested.*

*An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.*

***Solutions** should be sent to the address given on the inside front cover. Two copies suffice.*

*A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a Monthly problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

*For instructions about submitting solutions of Problems, which should be mailed before November 30, 1989, see the inside front cover. Please place the solver's name and mailing address on each (doubled-spaced) sheet. Include a self-addressed card or label if an acknowledgment is desired.*

## ELEMENTARY PROBLEMS

E 3331. *Proposed by Walter Rudin, University of Wisconsin, Madison.*

Exhibit a positive continuously differentiable function $f$ on $[0, \infty)$ such that

$$f'(x) \geqslant f^2(x) \qquad (0 \leqslant x < \infty),$$

or prove that no such function exists.

E 3332. *Proposed by John Mamer, University of California at Los Angeles, and Kenneth Schilling, University of Michigan-Flint.*

Let $X$ be a binomial random variable with parameters $n$ and $p$.

(a) Prove that

$$E(\min\{X, t\}) \geqslant t\{1 - (1 - p/t)^n\}$$

for $t = 1, 2, \ldots, n$.

(b)$*$ Is there an upper bound for $E(\min\{X, t\})$ better than $\min(np, t)$?

E 3333. *Proposed by Louis Comtet, Université de Paris-Sud, France.*

For $n = 1, 2, \ldots$, let $L_n$ denote the length of the curve $y = x^n$, $0 \leqslant x \leqslant 1$. Find an asymptotic formula for $2 - L_n$ as $n \to \infty$.

E 3334. *Proposed by Stanley Rabinowitz, Alliant Computer Systems Corporation, Littleton, MA.*

Consider the cubic curve $y = x^3 + ax^2 + bx + c$, where $a, b, c$ are real numbers with $a^2 - 3b > 3\sqrt{3}$. Prove that there are exactly two lines that are perpendicular

(normal) to the cubic at two points of intersection and that these two lines intersect at the point of inflection of the curve.

E 3335. *Proposed by D. E. Knuth, Stanford University, CA.*

Solve the recurrence

$$x_0 = a, \qquad x_1 = b, \qquad x_{n+2} = x_{n+1} + x_n/(n+1) \quad \text{for } n = 0,1,2,\ldots$$

both exactly (in terms of familiar functions of $n$) and asymptotically.

E 3336. *Proposed by Nick MacKinnon, Winchester, England.*

Suppose $P$ is a given polynomial with integer coefficients and degree greater than one. Let

$$W = \{ w : w \text{ is an irrational real number but } P(w) \text{ is rational}\}.$$

Prove that $W$ is dense on the real line.

## SOLUTIONS OF ELEMENTARY PROBLEMS

### A Double Infinite Sum for $|x|$

E 2982 [1983, 54]. *Proposed by R. L. Graham, Bell Telephone Laboratories and D. E. Knuth, Stanford University.*

Let $\|y\|$ denote the distance from (real) $y$ to the nearest integer. Evaluate the double infinite sum

$$\cdots + 4\|x/4\|^2 + 2\|x/2\|^2 + \|x\|^2 + \|2x\|^2/2 + \|4x\|^2/4 + \cdots .$$

*Solution prepared by the editors based on the solution of David Iny, Rensselaer Polytechnic Institute, Troy, NY.* The given sum is equal to $|x|$. We begin with the following lemma.

LEMMA. *Suppose $N$ is a given positive integer. If $m$ is a non-negative integer less than $2^N$, then*

$$\sum_{j=1}^{N} 2^j\|2^{-j}m\|^2 = m - 2^{-N}m^2.$$

*Proof.* The assertion of the lemma is true for $N = 1$. Assume that the assertion of the lemma is true for a particular value of $N$, and that we wish to prove it with $N$ replaced by $N + 1$. Suppose $M$ is a non-negative integer less than $2^{N+1}$. Then $M = \varepsilon \cdot 2^N + m$, where $\varepsilon$ is 0 or 1 and $0 \leqslant m < 2^N$. By the inductive assumption

$$\begin{aligned}\sum_{j=1}^{N+1} 2^j\|2^{-j}M\|^2 &= \sum_{j=1}^{N} 2^{-j}\|2^{-j}m\|^2 + 2^{N+1}\|2^{-N-1}M\|^2\\ &= m - 2^{-N}m^2 + 2^{N+1}\|2^{-N-1}M\|^2.\end{aligned}$$

If $\varepsilon = 0$, then $M = m < 2^N$, so that

$$2^{N+1}\|2^{-N-1}M\|^2 = 2^{-N-1}M^2$$

and

$$\sum_{j=1}^{N+1} 2^j\|2^{-j}M\|^2 = M - 2^{-N-1}M^2.$$

If $\varepsilon = 1$, then $2^N \leqslant M = m + 2^N < 2^{N+1}$, so that

$$2^{N+1}\|2^{-N-1}M\|^2 = 2^{N+1}(1 - 2^{-N-1}M)^2$$

and

$$\sum_{j=1}^{N+1} 2^j\|2^{-j}M\|^2 = (M - 2^N) - 2^{-N}(M - 2^N)^2 + 2^{N+1}(1 - 2^{-N-1}M)^2$$
$$= M - 2^{-N-1}M^2.$$

Thus the lemma is proved.

Now let

$$S(x) = \sum_{j=-\infty}^{\infty} 2^j\|2^{-j}x\|^2.$$

From the lemma we obtain

$$S(m) = \sum_{j=1}^{\infty} 2^j\|2^{-j}m\|^2 = m$$

for any nonnegative integer $m$. Since $S$ is an even function, we have $S(m) = |m|$ for any integer $m$. Further, if $m$ is any integer and $k$ is any positive integer,

$$S(2^{-k}m) = \sum_{j=-\infty}^{\infty} 2^j\|2^{-j-k}m\|^2$$
$$= 2^{-k}\sum_{j=-\infty}^{\infty} 2^{j+k}\|2^{-j-k}m\|^2$$
$$= 2^{-k}S(m) = |2^{-k}m|.$$

Thus $S(x) = |x|$ if $x$ is any rational number whose denominator is a power of 2.

Using the inequality $\|y\| \leqslant \min(1, |y|)$, we find that

$$2^j\|2^{-j}x\|^2 \leqslant \min(2^j, 2^{-j}|x|^2).$$

Thus the sum defining $S(x)$ is uniformly convergent for $x$ in any interval of the form $[-K, K]$. Hence $S$ is continuous and so $S(x) = |x|$ for all real $x$.

Solutions were received from 35 other readers and the proposers.

### A Lower Bound for the Ratio of the Median Sum to the Perimeter of a Triangle

E 3192 [1987, 181]. *Proposed by D. Fisher, Harvey Mudd College, and M. Martelli, Bryn Mawr College.*

Find the largest $k$ such that the following is true: If $\mathbf{u}, \mathbf{v}, \mathbf{w}$ are vectors in a normed space with $\mathbf{u} + \mathbf{v} + \mathbf{w} = 0$,

$$|\mathbf{u} - \mathbf{v}| + |\mathbf{v} - \mathbf{w}| + |\mathbf{w} - \mathbf{u}| \geqslant k(|\mathbf{u}| + |\mathbf{v}| + |\mathbf{w}|).$$

*Solution I by M. Falkowitz, Tel Aviv, Israel.* The answer is $k = 3/2$. Let $A$ be the set of numerical values assumed by the quotients

$$\frac{|\mathbf{u} - \mathbf{v}| + |\mathbf{v} - \mathbf{w}| + |\mathbf{w} - \mathbf{u}|}{|\mathbf{u}| + |\mathbf{v}| + |\mathbf{w}|},$$

where $\mathbf{u}, \mathbf{v}, \mathbf{w}$ sum to 0 but are not all 0. The desired $k$ is $\inf A$. We will show that $a \in A$ implies also $3/a \in A$. This yields $\inf A = 3/\sup A$. By the triangle inequality, $\sup A \leqslant 2$, achieved when one of the three vectors is zero; thus we can infer $k = 3/2$.

Given such $\mathbf{u}, \mathbf{v}, \mathbf{w}$, define $\mathbf{x} = \mathbf{u} - \mathbf{v}$, $\mathbf{y} = \mathbf{v} - \mathbf{w}$, $\mathbf{z} = \mathbf{w} - \mathbf{u}$. Then $\mathbf{x} + \mathbf{y} + \mathbf{z} = 0$ and $|\mathbf{x}| + |\mathbf{y}| + |\mathbf{z}| > 0$, but also $\mathbf{x} - \mathbf{y} = -3\mathbf{v}$, $\mathbf{y} - \mathbf{z} = -3\mathbf{w}$, $\mathbf{z} - \mathbf{x} = -3\mathbf{u}$. Thus

$$\frac{|\mathbf{x} - \mathbf{y}| + |\mathbf{y} - \mathbf{z}| + |\mathbf{z} - \mathbf{x}|}{|\mathbf{x}| + |\mathbf{y}| + |\mathbf{z}|} = 3 \cdot \frac{|\mathbf{u}| + |\mathbf{v}| + |\mathbf{w}|}{|\mathbf{u} - \mathbf{v}| + |\mathbf{v} - \mathbf{w}| + |\mathbf{w} - \mathbf{u}|},$$

which completes the proof.

*Solution II and generalization by John Henry Steelman, Indiana University of Pennsylvania.* For $n \geqslant 2$, we show that $n/2$ is the infimum of the values of $(\Sigma_{1 \leqslant i < j \leqslant n}|\mathbf{v}_i - \mathbf{v}_j|)/(\Sigma_{1 \leqslant i \leqslant n}|\mathbf{v}_i|)$ over sets $\mathbf{v}_1, \ldots, \mathbf{v}_n$ that sum to 0 but are not all 0.

Since $\Sigma\mathbf{v}_i = 0$, each $\mathbf{v}_i$ satisfies $n\mathbf{v}_i = \Sigma_{j \neq i}(\mathbf{v}_i - \mathbf{v}_j)$. Applying the triangle inequality and summing over $i$ yields $n\Sigma|\mathbf{v}_i| \leqslant 2\Sigma_{1 \leqslant i < j \leqslant n}|\mathbf{v}_i - \mathbf{v}_j|$. Hence the desired ratio is always at least $n/2$. This value is attained when $n - 1$ of the vectors are equal.

*Editorial comment.* Several readers mentioned the geometric interpretation of this result: in any normed space the lengths of the medians of a triangle sum to at least three-quarters of the perimeter of the triangle. Falkowitz provided a generalization for cyclic differences, namely, that the infimum of the values of $(|\mathbf{v}_1 - \mathbf{v}_2| + |\mathbf{v}_2 - \mathbf{v}_3| + \cdots + |\mathbf{v}_n - \mathbf{v}_1|)/\Sigma|\mathbf{v}_i|$ over sets of $n$ vectors summing to 0 but not all equaling 0 is $4n/(n^2 - 1)$ if $n$ is odd and $4/n$ if $n$ is even.

Also solved by K. A. Brown, Jr., J. Ferrer (Spain), S. Gudder, O. P. Lossers (The Netherlands), W. Margulies, I. Rosenholtz, the University of South Alabama Problem Group, and the proposers.

### An Elliptical Locus

E 3195 [1987, 300]. *Proposed by L. Kuipers, Sierre, Switzerland.*

Given triangle $ABC$, consider those inscribed ellipses touching $AB$ in $C_1$, $BC$ in $A_1$, and $CA$ in $B_1$, with $(AB_1)(BA_1) = (B_1C)(A_1C)$. Describe the locus of the centers of such ellipses.

*Solution by R. H. Eddy, Memorial University of Newfoundland, Canada.* If we denote by $(x, y, z)$ the barycentric coordinates of a point $P$ in the plane of the triangle of reference $ABC$, then $A_1 = (0, k, 1)$ and $B_1 = (1, 0, k)$, where $AB_1/B_1C = A_1C/BA_1 = k > 0$. The condition on $k$ stems from the fact that the given condition makes sense only when $A_1$ and $B_1$ divide the segments $BC$ and $CA$

internally, which further implies that the only such inscribed conics are ellipses. The equation of this family is given by

$$E_k\colon k^4x^2 + y^2 + k^2z^2 - 2k^2xy - 2kyz - 2k^3xz = 0.$$

The center of a conic is the pole of the line at infinity $x + y + z = 0$ with respect to that conic, hence the center of $E_k$ is given by the triple $(k + 1,\ k^2 + k,\ k^2 + 1)$. Eliminating $k$, we obtain the ellipse

$$F\colon x^2 + y^2 - xz - yz = 0.$$

Hence the required locus is the open elliptical arc determined by the intersection of $F$ with the interior of $ABC$. Note that the centroid $(1,1,1)$ is also a point in the locus; it is the center of the ellipse of maximum area that can be inscribed in the triangle $ABC$.

Also solved by L. E. Mattics and the proposer.

### Ten Points on a Conic

E 3232 [1987, 876]. *Proposed by Jordi Dou, Barcelona, Spain.*

Given lines $l_1, l_2, l_3, l_4, l_5$, and points $Q_1, Q_2, Q_3, Q_4, Q_5$, in the plane such that $Q_i$ does not lie in $l_i$ $(i = 1,2,3,4,5)$, prove that there exist points $P_i$, $R_i$ on line $l_i$ such that the angle $P_iQ_iR_i$ is a right angle $(i = 1,2,3,4,5)$ and such that the ten points $P_1, P_2, P_3, P_4, P_5, R_1, R_2, R_3, R_4, R_5$ lie on a conic.

*Solution by Richard L. Bishop, University of Illinois at Urbana-Champaign.* For a line $l$ and a point $Q$ not on $l$, we can make $l$ correspond to the pencil of lines through $Q$, viewed as a projective line. In this way a rotation by a right angle about $Q$ corresponds to a transformation of $l$ to itself such that a point and its image always subtend a right angle from $Q$. The transformation of $l$ is an elliptic involution, that is, a projective transformation with no fixed point and with square equal to the identity transformation. For example, if we take $Q$ to be $(0,1)$ and $l$ to be the $x$-axis, then the formula for the transformation of the $x$-axis is $x' = -1/x$. This example is canonical, up to projective coordinate changes.

LEMMA. *The condition that a real conic intersect a line in a pair of points exchanged by an elliptic involution is a linear restriction on the conic.*

*Proof.* By hypothesis we are given a line with an elliptic involution. Choose nonhomogeneous coordinates $(x, y)$ so that the line is given by $y = 0$ and the equation for the involution is $x' = -1/x$. The product of the $x$-coordinates of corresponding points is thus always $-1$.

A general conic has the form

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0,$$

so that the condition that the roots for $y = 0$ be points exchanged by the involution is given usually by $F/A$ = product of roots $= -1$, or the linear condition $F = -A$. The only exception is when $A = 0$, in which case $x = \infty$ is a root and the condition that $x = 0$ also be a root becomes $F = 0 = -A$ again.

*Note*. The linear relation $F = -A$ is also satisfied by $A = D = F = 0$, in which case the whole line is included in the (degenerate) conic. The conclusion of the lemma should still be regarded as satisfied.

COROLLARY. *Given* 5 *lines in the real projective plane, each with an elliptic involution, then there exist* 5 *pairs of corresponding points so that all* 10 *points are contained in a conic.*

*Proof.* 5 homogeneous linear equations with 6 unknowns always admit a nontrivial solution.

*Notes*. 1. It is not required that the lines nor their involutions be distinct. However, if two of the lines are the same but the involutions for them are different, then the solution conics must each contain at least 3 points on the line, so are all degenerate.

2. There are elliptic involutions which are not given by the effect of a rotation by a right angle about a point. For example, the exchange of any pair of conjugate axes of an ellipse (i.e., lines through the center which are parallel to the tangents at each others' intersections with the ellipse) is an elliptic involution of their pencil.

3. As a special case we have: Given a pentagon and a point not on any of its sides, there is a conic which intersects every side in pairs of points subtending a right angle from the given point.

*Editorial Comment*. The published problem statement is a generalization of the problem statement originally submitted by J. Dou, which further required that the vertices of the right angles be the opposite vertex of the pentagon formed by the 5 lines. The only correct solutions were those by Bishop and Dou, and the idea of using elliptic involutions was in Dou's solution, which was accompanied by a construction of the points $P_i$, $R_i$.

**Producing a Ranking From Recent Results**

E 3240 [1987, 996]. *Proposed by C. L. Mallows, AT&T Bell Laboratories, Murray Hill, NJ.*

The tennis players in a certain club compete in singles matches. After each match is played, a ranking of the players is computed according to the following rule. Starting with the most recent match and working back though time, use the match results to build up a partial order. Ignore any match that is inconsistent with more recent results. Stop when the partial order is a complete order.

Prove or disprove: a player cannot improve his position by intentionally losing a match.

*Solution I by the proposer*. False. With seven players, suppose the following results are obtained in the order given, where *ab* means that *a* beats *b*.

$$72\ 51\ 36\ 74\ 67\ 56\ 52\ 34\ 41\ 24\ 23\ 13.$$

The ranking after the last match is $5 > 6 > 7 > 2 > 4 > 1 > 3$, with 2 in fourth place. However, if 2 now loses to 1, the ranking becomes $5 > 1 > 2 > 3 > 6 > 7 > 4$, with 2 in third place.

*Solution II by Karl David, Wells College, Aurora, NY.* False. Suppose there are four players, that 1 plays 2 and 3 plays 4 in the first round, and then that round $i$ matches the winners in round $i-1$ followed by the losers in round $i-1$ (if necessary), until a complete order is established. Suppose further that the lesser-numbered player in any pair is better, except that 3 can beat 1. Again write $ab$ to mean "$a$ beats $b$." If the results in order are

12 34 31 24,

then the resulting ranking is $3 > 1 > 2 > 4$. If player 1 is aware of everyone's strength and throws the initial match, then the outcomes under the procedure and assumptions described are

21 34 23 14 12,

in which case the resulting ordering is $1 > 2 > 3 > 4$.

*Editorial comment.* Solution II involves assumptions about the behavior of other players. They could foil the efforts of player 1 by throwing their games. For example, after 21, player 3 can lose to 4 so that 21 43 42 gives current ordering $4 > 2 > 1$, $4 > 3$. If now 13, then 1 is left in third place, lower than previously, so 1 must throw another match to postpone the conclusion. We ask whether there is an example in the spirit of solution I with fewer than seven players, i.e. where the final match produces a complete ordering under either outcome, and one of the players involved receives a higher rank with a loss in that match than with a win.

No other solutions were received.

**Out of Bounds**

E 3243 [1988, 50]. *Proposed by William F. Trench, Trinity University, San Antonio, TX.*

Suppose that $g$ is constant in sign and locally Riemann-integrable on $[0, +\infty)$. Suppose $f$ is nonincreasing on $[0, +\infty)$ and $\int_0^\infty f(x)\,dx = +\infty$. Show that

$$\int_0^\infty |f(x)\cos x + g(x)\sin x|\,dx = \int_0^\infty |f(x)\sin x + g(x)\cos x|\,dx = +\infty.$$

*Solution by Adam Riese, Wright State University, Dayton, OH.* Let $[a, b]$ be a subinterval of $[0, 2\pi]$ with $\sin x \geqslant \frac{1}{2}$ and $g(x)\cot x \geqslant 0$ for all $x \in (a, b)$. Such an interval exists; in fact, if $g(x)$ is non-negative we may choose $(a, b) = (\pi/6, \pi/2)$, and if $g(x)$ is nonpositive we may choose $(a, b) = (\pi/2, 5\pi/6)$. With $f$ nonincreasing, we thus have

$$\int_0^\infty |f(x)\sin x + g(x)\cos x|\,dx \geqslant \sum_{k=0}^{\infty} \int_{a+2k\pi}^{b+2k\pi} f(x)/2\,dx \geqslant \frac{b-a}{4\pi}\int_{2\pi}^{\infty} f(x)\,dx = \infty.$$

The argument for $\int_0^\infty |f(x)\cos x + g(x)\sin x|\,dx$ is similar.

*Editorial Comment.* Kim McInturff notes that the conclusion fails if $f$ is allowed to be an extended real-valued function, more specifically, if $\lim_{x\to 0+} xf(x) = 1$. For example, take $f(x) = +\infty$ if $x = 0$, $f(x) = \csc x$ if $0 < x < \pi/2$, and $f(x) =$

0 if $x \geqslant \pi/2$, and put $g(x) = e^{-x}$. Then $f$ and $g$ satisfy the hypotheses, but

$$\int_0^\infty |f(x)\sin x + g(x)\cos x|\,dx = \int_0^{\pi/2}(1 + e^{-x}\cos x)\,dx + \int_{\pi/2}^\infty e^{-x}|\cos x|\,dx$$
$$= \frac{\pi}{2} + \int_0^\infty e^{-x}|\cos x|\,dx.$$

Other solvers: E. Bischof, B. Caccia, T. Hermann (Hungary), M. Hildebrand, I. E. Leonard (Canada), J.-M. Monier (France), and the proposer.

### A Class of Decreasing Functions

E 3244 [1988, 50]. *Proposed by R. S. Rodriguez and H. Sherwood, University of Central Florida, Orlando.*

Suppose $a_0, a_1, \ldots, a_k$ are integers such that $0 < a_0 < a_1 < \cdots < a_k$. Put

$$f(x) = \frac{1}{(1+x)^{a_0}} + \frac{x}{(1+x)^{a_1}} + \frac{x^2}{(1+x)^{a_2}} + \cdots + \frac{x^k}{(1+x)^{a_k}}.$$

The function $f$ is easily seen to be strictly decreasing on $(0, +\infty)$ when each $a_i = i + 1$. Is $f$ strictly decreasing for every permissible choice of $a_0, a_1, \ldots, a_k$?

*Solution by B. S. Bertram and O. G. Ruehr, Michigan Technological University, Houghton, MI.* The answer is "yes." To see this, we compute

$$f'(x) = \frac{-a_k x^k}{(1+x)^{1+a_k}} - \sum_{j=0}^{k-1} x^j \left\{ \frac{a_j}{(1+x)^{1+a_j}} - \frac{j+1}{(1+x)^{a_{j+1}}} \right\}$$
$$= \frac{-a_k x^k}{(1+x)^{1+a_k}} - \sum_{j=0}^{k-1} \frac{x^j}{(1+x)^{1+a_j}}$$
$$\times \left\{ a_j - j - 1 + (j+1)\left(1 - \frac{1}{(1+x)^{p_j}}\right) \right\},$$

where $p_j = a_{j+1} - a_j - 1$. Since $a_j \geqslant j + 1$ and $p_j \geqslant 0$ for all $j$, $f'(x)$ is negative for $x > 0$.

Note that simply requiring $a_i \geqslant i + 1$ and omitting ordering restrictions on the $a_i$'s makes the assertion false. For example, take, $k = 1$, $a_0 = a$, and $a_1 = 2$. Then

$$f'\left(\frac{1}{2}\right) = -a\left(\frac{2}{3}\right)^{a+1} + \frac{4}{27}$$

is clearly positive for $a$ sufficiently large ($a = 10$ will do).

Other solvers: Nicolas K. Artemiadis (Greece), Stephen M. Gagola, T. Hermann (Hungary), Ha-Seo Ki (South Korea), L. Kuipers (Switzerland), K. McInturff, G. Parker (Scotland), J. T. Poole, J. Riley, E. M. Sanchez (Columbia), J. H. Steelman, and the proposers. One incorrect and one incomplete solution were received.

**Matrices Displaying Their Eigenvalues**

E 3245 [1988, 51]. *Proposed by Bruce A. Reznick and Lee A. Rubel, University of Illinois, Urbana.*

(a) For what positive integers $n$ does there exist an $n$ by $n$ matrix $A$ over $\mathbb{C}$ having the following three properties:

(i) $n^2 - n$ of the entries are zero,

(ii) there are $n$ distinct nonzero entries $r_1, r_2, \dots, r_n$, none of which lies on the main diagonal,

(iii) the eigenvalues of $A$ are $r_1, r_2, \dots, r_n$?

(b) Same as (a) except that all the entries are required to be real.

*Solution by David Callan, University of Bridgeport, Bridgeport, CT.* Such a matrix exists if and only if $n \geqslant 3$, but never with all entries real.

(a) It is straightforward to check that there is no solution for $n \leqslant 2$. For $n \geqslant 3$, suppose that $A$ has its nonzero entries $r_i$ in the positions $(\sigma(i), i)$, for some fixed-point-free permutation $\sigma$. Let $\sigma$ be a product of disjoint cycles $\sigma_1, \dots, \sigma_l$ on the partition $T_1 \cup \cdots \cup T_l$ of $\{1, \dots, n\}$, with $t_j = |T_j|$. Then the characteristic polynomial of $A$ is

$$p(\lambda) = \prod_{j=1}^{l} \left( \lambda^{t_j} - \prod_{k \in T_j} r_k \right),$$

which is immediate when $\sigma$ is a single cycle and follows otherwise by simultaneous row and column permutations to place the elements of each cycle in a diagonal block.

For $n \geqslant 3$, we must find some fixed-point free $\sigma$ and nonzero $r_i$'s such that $p(\lambda) = \prod_{i=1}^{n}(\lambda - r_i)$. Let $\omega$ be a primitive $n$th root of unity. If $n$ is odd, use a single cycle $\sigma = (1, 2, \dots, n)$ and let $r_i = \omega^i$. If $n = 2m$ is even, use two cycles $\sigma = (1, 2, \dots, m)(m+1, m+2, \dots, 2m)$ and let $r_i = \omega^{2i-1}$ for $1 \leqslant i \leqslant m$ and $r_i = \omega^{2i}$ for $m + 1 \leqslant i \leqslant 2m$.

(b) If the roots of $p(\lambda)$ are all real, then each factor $\lambda^{t_j} - \prod_{k \in T_j} r_k$ has real roots, so $t_j \leqslant 2$, and in fact $t_j = 2$ because $\sigma$ has no 1-cycles. Now $A^2$ is a diagonal matrix, and its eigenvalues, which appear on its diagonal, are the squares of the eigenvalues of $A$. Because $\sigma$ consists of disjoint two-cycles, the entries on the diagonal of $A^2$ are also the products of a matched pair of off-diagonal entries of $A$. If $r = \max\{|r_i|\}$, this means $r^2 = r_i r_j$, which is possible only if $r_i = r_j = \pm r$, contradicting the distinctness of the eigenvalues.

If we drop the distinctness requirement, this argument successively shows that all matched pairs of off-diagonal entries are equal. Since the factors $\lambda^2 - \prod_{k \in T_j} r_k$ generate roots in oppositely-signed pairs, dropping the distinctness requirement allows real solutions only when $n$ is divisible by 4, in which case the nonzero entries come in groups of the form $r, r, -r, -r$.

Also solved by J. T. Bruening, J. Ferrer (Spain), S. M. Gagola, Jr., G. N. Lewis, O. P. Lossers (The Netherlands), K. McInturff, J.-M. Monier (France), A. Pedersen (Denmark), J. H. Steelman, W. P. Wardlaw, E. C. Weinberg, and the proposers. Several incorrect solutions were received.

## ADVANCED PROBLEMS

6604. *Proposed by Michael Reid (student), Harvard College, Cambridge, MA.*

For which primes $p > 3$ is it true that

$$\sum_{k=0}^{p-1} \frac{(3k)!}{(k!)^3}(-6)^{3p-3k} \equiv 0 \pmod{p}?$$

6605. *Proposed by E. Ehrhart, Université de Strasbourg, France.*

We use the term "lattice point" to mean a point with integer coordinates.

If $k$ is a positive integer, Schinzel [*Enseignement Math.* (2) 4 (1958) 71–72] proved that the circle

$$\left(X - \frac{1}{3}\right)^2 + Y^2 = (5^k/3)^2$$

passes through exactly $2k + 1$ lattice points; clearly the two coordinates of any one of these $2k + 1$ lattice points are of like parity. Thus by making the substitution $X = x + y$, $Y = x - y$ we see that the smaller circle

$$\left(x - \frac{1}{6}\right)^2 + \left(y - \frac{1}{6}\right)^2 = (5^k\sqrt{2}/6)^2 \qquad (*)$$

also passes through exactly $2k + 1$ lattice points.

(i) Show that when $k = 1$ no circle smaller than $(*)$ passes through exactly three lattice points.

(ii) Show that when $k = 2$ no circle smaller than $(*)$ passes through exactly five lattice points.

(iii) Show that when $2k + 1$ is composite there is a circle smaller than $(*)$ which passes through exactly $2k + 1$ lattice points.

## SOLUTIONS OF ADVANCED PROBLEMS

### Nonisomorphic Fields

6548 [1987, 559]. *Proposed by Lee A. Rubel, University of Illinois at Urbana-Champaign.*

Are there uncountably many nonisomorphic fields $F_\alpha$, where $\alpha$ runs over some uncountable index set $A$, such that no two of the fields $F_\alpha$ are isomorphic but all of the additive groups $(F_\alpha, +)$ are isomorphic and all of the multiplicative groups $(F_\alpha, \cdot)$ are isomorphic. Cf. 6489 [1985, 148; 1986, 744].

*Solution by R. Schutt, Milwaukee, WI.* There are in fact uncountably many such fields contained in $\mathbb{R}$, the field of real numbers. Let $P$ denote the positive reals, and $E$ a maximal algebraically independent subset of $\mathbb{R}$. By replacing the appropriate elements of $E$ by their negatives, we may assume $E \subseteq P$. For each subset $A$ of $E$, let $F_A$ denote the algebraic closure of $A$ in $\mathbb{R}$. Then $F_A$ is a field (in fact a real closed

field) and $P_A$, defined by

$$P_A = F_A \cap P,$$

is characterized as the set of nonzero elements of $F_A$ that have a square root. Hence $P_A$ is preserved by every homomorphism. Thus if $A$ and $B$ are subsets of $E$ and $\theta$ is a homomorphism from $F_A$ to $F_B$, we have $\theta(P_A) \subseteq P_B$.

LEMMA. *If $F$ is a subfield of $\mathbb{R}$ and $f$ is an additive map of $S = F \cap P$ into $P$, then*

$$f(r) = rf(1)$$

*for all $r \in F \cap P$.*

*Proof.* If $a < b$ are both in $S$ then so is $b - a$, and

$$f(b) = f(b - a + a) = f(b - a) + f(a) > f(a).$$

Thus $f$ is increasing. Also, it is clear that

$$f(q) = qf(1)$$

for every positive rational $q$. Now suppose

$$f(x) > xf(1) \tag{*}$$

for some $x \in S$. Choose $k > 1$ real so that $f(x) = kxf(1)$, and a positive rational $q$ such that

$$x < q < kx.$$

Then

$$f(x) < f(q) = qf(1) < kxf(1) = f(x),$$

a contradiction. If (*) is replaced by the opposite inequality a similar contradiction ensues; hence, the result of the lemma follows.

Next, since $\theta(1) = 1$ for the above homomorphism, we have $\theta(r) = r$ for any $r \in P_A$. Thus there is a homomorphism from $F_A$ into $F_B$ if and only if $A \subseteq B$. It follows that $F_A$ and $F_B$ are not isomorphic if $A \neq B$.

Let $G$ denote either the additive group of $F_A$ or the multiplicative group of $F_A \cap P$. $G$ is torsion free and divisible (e.g., each element of $F_A \cap P$ has a unique $n$th root for each positive integer $n$). Hence, $G$ may be regarded as a vector space over the field $\mathbb{Q}$ of rationals. Moreover, $A$ is an independent subset of this space and can be extended to a basis. Let $|A|$ denote the cardinality of $A$. If $A$ is infinite, $|A| = |F_A| = |G|$ and, hence, the dimension of this space equals $|A|$. Thus, if $A$ is infinite, $G$ is isomorphic to the direct sum of $|A|$ copies of $\mathbb{Q}^+$, the additive group of $\mathbb{Q}$. Clearly, the multiplicative group of $F_A$ is isomorphic to the direct product of $G$ and the group of order 2. Thus $F_A$ and $F_B$ have isomorphic additive and isomorphic multiplicative groups if $A$ and $B$ are infinite and $|A| = |B|$.

The following conclusions now follow directly from the above. Let $c$ denote the cardinality of the continuum. Note that there are $c$ denumerable subsets of $E$ and $2^c$ subsets of cardinality $c$. Thus

(1) There is a set of $c$ denumerable nonisomorphic subfields of $\mathbb{R}$, all with isomorphic additive and multiplicative groups, and

(2) There is a set of $2^c$ nonisomorphic subfields of $\mathbb{R}$, all with isomorphic additive and multiplicative groups. (In fact they are isomorphic to those of $\mathbb{R}$ itself.)

*Editorial Comment*. The proposer's solution appears quite different. For $0 < \rho < 1$ define $E(\rho)$ to be the ring of entire functions of order strictly less than $\rho$; in other words $E(\rho)$ is the set of entire functions satisfying

$$|f(z)| \leqslant \alpha e^{\beta |z|^{\rho - \epsilon}} \text{ for all } z,$$

where $\alpha$, $\beta$, $\epsilon$ are positive constants depending on $f$. Next let $K(\rho)$ be the quotient field of $E(\rho)$; this is a field of meromorphic functions. That different values of $\rho$ yield nonisomorphic fields is essentially a result of James J. Kelleher, On isomorphisms of meromorphic function fields, *Canadian J. Math.*, 20 (1968) 1230–1241. Now all $K(\rho)$ are vector spaces over $\mathbb{C}$ of the same dimension, and hence have the same additive group. With the aid of the Weierstrass factorization theorem and an appropriate "stretching" of the plane, it can also be shown that the multiplicative groups are isomorphic.

No other solutions were received.

### "Almost Abelian" Groups

6553 [1987; 659]. *Proposed by Ian D. Macdonald, Menstrie, Scotland.*

Suppose that $G$ is a group with generators $a, b, c, d$, where $abcd = cdab$, $ab = ba$, $ac = ca$, $bd = db$, $cd = dc$. Prove that the commutator subgroup of $G$ is cyclic and contained in the center of $G$, but show that $G$ is not necessarily abelian.

*Combined solution by Thomas Jager, Calvin College, Grand Rapids, MI, and Lou Shapiro, Howard University, Washington, D.C.* From the given relation $abcd = cdab$ and the fact that $b$ and $c$ each commute with both $a$ and $d$, we have $(bc)(ad) = (cb)(da)$. Thus

$$g = b^{-1}c^{-1}bc = dad^{-1}a^{-1}$$

belongs to $Z(G)$, the center of $G$; in fact, since $a$ commutes with both $b$ and $c$, we have $ag = ga$, and similarly $bg = gb$, $cg = gc$, $dg = gd$. Further, since $ad = g^{-1}(da)$ and $bc = (cb)g$ and since the only non-commuting pairs of generators of $G$ are the pair $a, d$ and the pair $b, c$, it follows that the group $G/\langle g \rangle$ is abelian. Hence the commutator subgroup $G'$ satisfies

$$G' \subseteq \langle g \rangle \subseteq Z(G).$$

Now $\langle g \rangle \subseteq G'$, so $G' = \langle g \rangle$. Thus the first assertion of the problem is established.

For a nonabelian example, let $G$ be the set of all $4 \times 4$ matrices over the integers modulo 2 of the form

$$\begin{pmatrix} 1 & p & q & r \\ 0 & 1 & 0 & s \\ 0 & 0 & 1 & t \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Choose $a$, $b$, $c$, $d$, and $m$ to be certain elements of $G$ for which exactly one off-diagonal element is nonzero, namely $p$ for $a$, $q$ for $b$, $r$ for $m$, $s$ for $d$, and $t$ for

$c$. Then $ad \neq da$, $bc \neq cb$, but the postulated equalities are valid. This establishes the second assertion of the problem; we remark that in this group of order 32 the commutator subgroup consists of precisely the two elements 1 and $m$.

*Other solvers*: Roger Alperin, Charles C. Edmunds (Canada), Jesús Ferrer (Spain), A. M. Gaglione and Dennis Spellman (jointly), Stephen M. Gagola, Jr., Michael Josephy (Costa Rica), O. P. Lossers (The Netherlands), Derek J. S. Robinson, John Henry Steelman, Douglas B. Tyler, Gary L. Walls, and the proposer.

# REVIEWS

EDITED BY JOSEPH KONHAUSER
*Department of Mathematics, Macalester College, St. Paul, MN 55105*

*Mathematics with Applications, Fourth Edition.* By Margaret L. Lial and Charles D. Miller. Scott, Foresman and Company, Glenview, Illinois, 1987. x + 740 pp.

CLINTON J. OXENRIDER
*Department of Mathematics and Computer Science, Bloomsburg University, Bloomsburg, PA 17815*

Here is a textbook that covers, in two semesters, those topics traditionally found in college algebra, linear algebra, probability and statistics, and calculus. We, those students brought up in the traditional mathematics curriculum, all know that these topics require at least four or five semesters. Whenever I see a textbook or a course that purports to telescope material in this manner the following questions come to mind. Why should the material be packaged in this manner when texts already exist presenting the material in more discrete packages? How can this much material be intelligibly covered in such a short time? What is the quality of the mathematics in such a course? How does the existence of such courses affect the mathematics curriculum and mathematics in general?

The first question is easily answered. Many programs exist that require mathematics to service the highly specialized content of the program in as short a time as possible. This is usually true of the business curriculum, which offers such courses as business and economic mathematics, and the social and political sciences curricula.

To cover this much material in so short a time is typically a function of the instructor's ability to develop an intuitively clear presentation of each topic on the first trial. If the student intuitively believes the presentation, there is no need for rigor and logical correctness. For those students that do not understand on the first trial, there are usually enough drill exercises to convince them.

The quality of the mathematics in such a course is a function of the textbook and, especially, the instructor's abilities. Unfortunately, too often the instructor will concentrate only on those topics considered to be of importance to the specialization, or, worse yet, will present the entire course in a "monkey-see, monkey-do" fashion. Also, in many schools such courses are not taught by the mathematics department or anyone with credentials in mathematics. The result is classes with high mathematical competencies ranging down to classes in mathematical shock. Certainly the quality of mathematics suffers in an environment when the question "How?" is more important than the question "Why?"

Such courses do indeed make a significant audience of students aware of the practical uses of mathematics for low level applications. For many it is a strong motivational tool to do further work in mathematics when they discover the mathematics they learned in such courses is not adequate to understand and develop more complicated models. I have personally noted that more such students are enrolling in mathematics courses each semester even though these courses are not required in their programs. When the course is properly taught I believe it has a definite positive effect on the mathematical maturity of its audience.

The Lial and Miller text is adaptable for a two-semester course in finite mathematics and calculus, a one-semester course in finite mathematics, a one-semester course in calculus, or a one-semester course in college algebra with applications. It is a more than adequate textbook for any of these options. The fourth edition contains realistic applications and topical case studies. It has more than 400 examples, 3900 drill exercises and 1300 applied exercises. I did a cursory count of the exercises that were listed by area of application and found 344 for management, 53 for social sciences, 211 for natural sciences, 2 for political science, 1 for biology and 6 for physical science (a somewhat biased distribution). The instructor's guide contains test items (with answers) and BASIC programs for Leontief models, Markov chains, and linear programming. I found the case studies to be of particular interest.

While this textbook is marketed for two-year colleges and college programs with strong emphasis in business, social science and political science programs, I feel it would be an excellent textbook for the high school curriculum. Each year many good high school students, for whatever reason, are eliminated from the usual sequence of algebra, geometry, trigonometry, calculus. The system calls it natural attrition, but too many of those students that successfully complete algebra never take another mathematics course. Also many smaller high schools cannot justify the expense of carrying a calculus course for the senior year. A one-year course based on this textbook would help correct both situations. The modest prerequisite of algebra would reclaim those students that dropped out of the mathematics curriculum after the beginning algebra courses. The intuitive approach and the realistic and timely applications should remotivate these students. They stand a chance of improving their mathematical skills, developing expertise in an applied area, becoming knowledgeable users of computer software, etc. This alternative seems far superior to any that I have seen suggested thus far.

*Sphere Packings, Lattices and Groups.* By J. H. Conway and N. J. A. Sloane. Springer-Verlag, New York, Berlin, Heidelberg, 1988. xxvii + 663 pp.

H. S. M. Coxeter
*Department of Mathematics, University of Toronto, Ontario, Canada M5S 1A1*

In old war memorials one sometimes sees a square pyramidal heap of cannon balls: one at the top resting on four, these resting on nine in the third layer, ..., and $n^2$ in the $n$th layer. Clearly, an interior ball touches 12 others: 4 in its own layer, 4 above and 4 below. Their 12 centers, and similarly their 12 points of contact with the original ball, are the vertices of a cuboctahedron (Figure 1). The balls, so arranged, are said to be in *cubic close-packing*. When a model of such a pyramidal heap, with the balls glued together, is turned over so that one of the four triangular side-faces becomes horizontal, the same arrangement of balls is seen to be distributed into hexagonal layers, so that the 12 neighbors of an interior ball now consist of 6 in its own layer, 3 above and 3 below. (The cuboctahedron is now standing on a triangular face instead of a square face.) This cubic close-packing of balls, continued indefinitely in all directions, is called a *lattice* packing because the

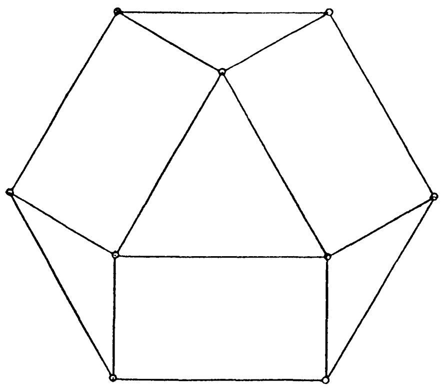


FIG. 1. The hexagon $e\alpha_2$ and the cuboctahedron $e\alpha_3$.

centers form a lattice: the position vectors of all the centers, as seen from any one of them, form an additive group of vectors.

When two consecutive layers are given, each ball in the upper layer is surrounded by 6 "deep holes," only 3 of which can accommodate balls in the next higher layer. In fact, this next layer can be filled in two different ways: one as in the lattice packing, and the other so that each ball is exactly above one in the bottom layer. If the latter rearrangement is continued, so that every two alternate layers "agree," we have *hexagonal close-packing* [**15**, p. 170; **1**, p. 150]. Since there are two possible positions for each new layer, hexagonal close-packing is one of infinitely many *nonlattice* packings, all equally dense (or 'economical') and having the same *kissing number* 12. But it is only in the lattice packing that the 12 points of contact form a cuboctahedron and, by reciprocation, the *Voronoi cell* (bounded by the 12 tangent planes) is a rhombic dodecahedron.

All these ideas were already understood around 1600 by Kepler and his English friend Thomas Harriot [**11**, pp. 14–17, 29, 52; **12**, pp. 23–24, 60–62]. Kepler described the Voronoi cell in terms of a pomegranate, as the shape of its loculi, that is, the juicy morsels that surround the seeds.

If the balls have radius 1, the volume of the rhombic dodecahedron, being twice that of a cube of edge $\sqrt{2}$, is $4\sqrt{2}$. Accordingly, the *center density* (that is, the number of centers per unit volume) is $1/4\sqrt{2}$.

This book by Conway and Sloane (with additional contributions by E. Bannai, J. Leech, S. P. Norton, A. M. Odlyzko, R. A. Parker, L. Queen and B. B. Venkov) describes the enormous progress that has been made during the last few decades. Yet it is remarkable that some apparently simple questions still remain unanswered. In the famous words of Ambrose Rogers, quoted on page 3, "many mathematicians believe, and all physicists know" that these close-packings, with center density $2^{-5/2}$, are denser than any other arrangement of congruent balls filling Euclidean 3-space; but this fact has never been rigorously proved! In fact, the densest ball packing in $n$ dimensions is known only when $n = 1$ or 2.

A set of points evenly spaced along a line, the centers of a hexagonal close-packing of circular disks or "2-balls," and the centers of the lattice close-packing of "3-balls," are the instances $\Lambda_1$, $\Lambda_2$ and $\Lambda_3$ of an $n$-dimensional *laminated* lattice

$\Lambda_n$. The $n$-balls serve as "equators" for the initial layer of $(n + 1)$-balls in the analogous lattice $\Lambda_{n+1}$ (see page 143). The points of contact of one $(n + 1)$-ball with its neighbors are the vertices of a polytope $\Pi_{n+1}$ (the cuboctahedron when $n = 2$) having $\Pi_n$ for its "equatorial" section [**2**, p. 414]. There are also "tropical" sections $\pi_n$ containing other vertices of $\Pi_{n+1}$. In most cases, all the vertices are distributed in the three parallel sections $\pi_n$, $\Pi_n$, $\pi_n$. (When $n = 2$, they are "triangle, hexagon, triangle.") But occasionally these are preceded and followed by one vertex at the "north pole" and its antipodes at the "south pole." Since every layer of $\Lambda_{n+1}$ is a $\Lambda_n$, $\pi_n$ occurs as a cell of the $n$-dimensional honeycomb whose vertices are the points of $\Lambda_n$. We call $\Pi_n$ the *vertex figure* of that honeycomb.

When $n = 2$, we have the lattice $\Lambda_2$ whose points form the tessellation $\{3,6\}$ of equilateral triangles, 6 around each vertex. If we take its edges to have length 2, the "2-balls" are circles of unit radius with their centers at the vertices. The regions of the plane outside the circles are called *holes* and there is clearly one hole inside each triangle. When the circles are blown up into spheres or "3-balls" for the construction of $\Lambda_3$, the holes are ready to accommodate the next higher layer of balls. The centers in this next layer form another $\{3,6\}$, one of whose triangles is the top face $\pi_2$ of the cuboctahedron $\Pi_3$. Its equatorial section is the hexagon $\Pi_2$ (one tile of the tessellation $\{6,3\}$ dual to $\{3,6\}$). The vertices of this hexagon are in 6 holes, only 3 of which can lie just below the centers of balls in the higher layer. Thus the possibility of changing cubic close-packing into hexagonal arises from the fact that the hexagon has two inscribed equilateral triangles.

The points of the lattice $\Lambda_3$ are the vertices of a honeycomb of octahedra $\beta_3$ and tetrahedra $\alpha_3$. This honeycomb [**1**, p. 148] is denoted by $h\delta_4$ because it has half the vertices of the ordinary cubic lattice $\delta_4$, of edge $\sqrt{2}$. (The more usual designation "face-centered cubic lattice" or *fcc* refers to the faces of a $\delta_4$ of edge $2\sqrt{2}$.)

Cubic close-packing admits, inside each octahedron or tetrahedron of $h\delta_4$, a "deep" or "shallow" hole formed by a cluster of six or four balls, respectively [**14**, p. 180]. When these 3-balls are blown up into 4-balls for the construction of $\Lambda_4$, the deep holes are ready to support the next higher layer of 4-balls ("higher" along the fourth dimension). The centers in this next layer form another $h\delta_4$, one of whose octahedra is the "top" facet $\pi_3$ of the polytope $\Pi_4$. This polytope is Ludwig Schläfli's regular 24-cell $\{3,4,3\}$, which may alternatively be called $t_1\beta_4$ because its 24 vertices are the mid-edge points of a 16-cell $\beta_4 = \{3,3,4\}$ [**5**, p. 575]. (Analogously, the cuboctahedron could be called $t_1\ \beta_3$ because its 12 vertices are the mid-edge points of an octahedron $\beta_3 = \{3,4\}$.) The vertices of $\Pi_4 = \{3,4,3\}$ are distributed in layers as $6 + 12 + 6$ [**3**, p. 298]. Its equatorial section is the cuboctahedron $\Pi_3$ whose 6 square faces surround 6 deep holes, all just "below" the vertices of the octahedron $\pi_3$. Thus, in contrast to the 3-dimensional case, the lattice $\Lambda_4$ provides the only known packing with kissing number 24 and center density $1/8$: there is no known equally dense nonlattice packing.

This $\Lambda_4$ is denoted by $D_4$, to distinguish it from a slightly less dense packing $A_4$ (with layers farther apart) which can be derived from $\Lambda_3(= A_3 = D_3)$ by using the shallow holes (of the cubic close-packing) instead of the deep holes, so that $\pi_3$ is a tetrahedron while $\Pi_3$ remains a cuboctahedron, and $\Pi_4$ is Alicia Boole Stott's "expanded simplex" $e\alpha_4$, whose facets consist of 10 tetrahedra and 20 triangular prisms [**5**, p. 574]. Its 20 vertices are distributed in layers as $4 + 12 + 4$. The 8 triangular faces of $\Pi_3$ surround 8 shallow holes, only 4 of which can be just below

the centers of balls in the higher layer so as to yield a tetrahedral facet of the higher $\Lambda_3$. Just as in the transition from cubic to hexagonal close-packing, we can rotate this higher $\Lambda_3$ so as to use the 4 points just above the remaining 4 of the 8 shallow holes, and so to construct a nonlattice packing which has the same kissing number 20, and the same center density $1/4\sqrt{5}$, as the lattice packing $A_4$. This nonlattice packing may thus be said to arise from the fact that the cube has two inscribed tetrahedra [**11**, p. 11; **12**, p. 50].

$A_4$ belongs to a sequence of lattice packings $A_n$ ($= \Lambda_n$ when $n < 4$) with kissing number $n(n+1)$ and center density $2^{-n/2}(n+1)^{-1/2}$. The points of $A_n$ are the vertices of P. H. Schoute's honeycomb $\alpha_n h$ [**8**, p. 152] whose cells are the regular simplex $\alpha_n$ and all its principal truncations $\mathrm{t}_\mu \alpha_n$. (The vertices of $\mathrm{t}_\mu \alpha_n$ are the centers of the $\mu$-faces of $\alpha_n$.) Each cell $\alpha_n$ belongs to a cluster of $n+1$ mutually tangent $n$-balls surrounding a shallow hole. When the $n$-balls are blown up into $(n+1)$-balls for the construction of $A_{n+1}$, such holes are ready to support the next higher $\alpha_n h$, one of whose $\alpha_n$'s is the top facet $\pi_n$ of the polytope $\Pi_{n+1}$. This $\Pi_{n+1}$ is an expanded simplex $e\alpha_{n+1}$, whose facets are simplexes $\alpha_n$ and generalized prisms; more precisely, they are Cartesian products $\alpha_m \times \alpha_{n-m}$ of simplexes [**4**, p. 128].

FIGURE 1 illustrates both the cases $n = 1$ and $n = 2$: the hexagon $e\alpha_2$ is an *expansion* [**1**, pp. 139–140] of the triangle $\alpha_2$, and the cuboctahedron $e\alpha_3$ (with faces $\alpha_2$ and $\alpha_1 \times \alpha_1$) is an expansion of the tetrahedron $\alpha_3$. (Thus $e\alpha_3 = \mathrm{t}_1\beta_3$ and $\alpha_3 h = \mathrm{h}\delta_4$.) The $(n+1)(n+2)$ vertices of $e\alpha_{n+1}$ are distributed in layers as

$$(n+1) + n(n+1) + (n+1).$$

The $2(n+1)$ facets $\alpha_n$ of $e\alpha_{n+1}$ surround this number of shallow holes, only $n+1$ of which can lie just below the centers of balls in the higher layer so as to yield a simplicial facet $\alpha_n$ of the higher $\alpha_n h$. Thus, along with the lattice packing $A_{n+1}$, there is an equally dense nonlattice packing.

Similarly, $D_4$ belongs to a sequence of lattice packings $D_n$ with kissing number $2n(n-1)$ and center density $2^{-1-n/2}$. The points of $D_n$ are the vertices of a honeycomb $\mathrm{h}\delta_{n+1}$ of cross polytopes $\beta_n$ (reciprocal to the $n$-cube $\gamma_n$) and hemi-cubes $\mathrm{h}\gamma_n$ [**3**, p. 156]. (The $2^{n-1}$ vertices of $\mathrm{h}\gamma_n$ are alternate vertices of $\gamma_n$.) Inside these cells ($\beta_n$ and $\mathrm{h}\gamma_n$) of $\mathrm{h}\delta_{n+1}$ there are holes bounded by clusters of $2n$ or $2^{n-1}$ $n$-balls. When these balls are blown up into $(n+1)$-balls for the construction of $D_{n+1}$, the holes of either kind are ready to support the next higher $D_n$. One of the cells of this $\mathrm{h}\delta_{n+1}$ is the top facet $\pi_n$ of the polytope $\Pi_{n+1}$, whose equatorial section is $\mathrm{t}_1\beta_n$. The "truncated cross polytope" $\mathrm{t}_1\beta_n$ has $2n$ facets $\beta_{n-1}$ (corresponding to the vertices of $\beta_n$) and $2^n$ facets $\mathrm{t}_1\alpha_{n-1}$ (truncations of the simplicial facets $\alpha_{n-1}$ of $\beta_n$).

If $\pi_n$ is $\beta_n$, $\Pi_{n+1}$ is $\mathrm{t}_1\beta_{n+1}$, whose $2n(n+1)$ vertices are distributed in layers as

$$2n + 2n(n-1) + 2n.$$

Its equatorial section is the $\mathrm{t}_1\beta_n$ whose $2n$ facets $\beta_{n-1}$ and $2^n$ facets $\mathrm{t}_1\alpha_{n-1}$ surround the two kinds of holes, which are distinct if $n \neq 4$. The $2n$ holes of the former kind are just below the vertices of $\pi_n = \beta_n$, which is one of the cells of the higher $\mathrm{h}\delta_{n+1}$. We thus obtain the honeycomb $\mathrm{h}\delta_{n+2}$ and the lattice $D_{n+1}$.

The case when $n = 4$ is exceptional because the polytopes $\mathrm{h}\gamma_4, \mathrm{t}_1\beta_4$ and the honeycomb $\mathrm{h}\delta_5$ are *regular* [**3**, pp. 148, 156; **5**, p. 575]:

$$\mathrm{h}\gamma_4 = \beta_4, \qquad t_1\beta_4 = \{3,4,3\}, \qquad \mathrm{h}\delta_5 = \{3,3,4,3\}.$$

These coincidences eliminate the distinction between the two kinds of holes. Since $\{3,4,3\}$ has three inscribed 16-cells $\beta_4$, the 8 balls whose centers belong to $\pi_4 = \beta_4$ can be shifted to either of two new locations, so there are now non-lattice packings having the same kissing number 40 and the same center density $2^{-7/2}$ as $D_5$ (see page 144).

Returning to the discussion of deriving an $(n+1)$-dimensional lattice from $D_n$, we naturally ask what happens if $\pi_n$ is not $\beta_n$ but $\mathrm{h}\gamma_n$. When $n = 4$ (as we have just seen) this makes no difference. When $n = 3$ we have $D_3 = A_3$ and $\mathrm{h}\gamma_3 = \alpha_3$, so the 4-dimensional lattice is merely $A_4$. The excitement begins when $n = 5$, so that $\pi_5 = \mathrm{h}\gamma_5$ is one of the facets of a different polytope $\Pi_6$ whose equatorial section $\Pi_5$ is again $\mathrm{t}_1\,\beta_5$. Since the vertices of $\Pi_6$ are distributed in layers as $16 + 40 + 16$, their number is 72. In fact, this $\Pi_4$ is the $\mathrm{V}_{72}$ of Elte [**9**, pp. 104–108; **4**, p. 131] which was later recognized as the member $1_{22}$ of a remarkable family $p_{qr}$ of $(p+q+r+1)$-dimensional polytopes, and $(p+q+r)$-dimensional honeycombs, which occur whenever

$$pqr \leqslant p + q + r + 2$$

(with equality for honeycombs). To be precise, $p_{qr}$ $(= p_{rq})$ is the polytope or honeycomb (with triangular 2-faces) whose vertex figure is $(p-1)_{qr}$, the vertex figure of $1_{qr}$ being the truncated simplex

$$0_{qr} = \begin{Bmatrix} 3^q \\ 3^r \end{Bmatrix} = \mathrm{t}_q\alpha_{q+r+1} = \mathrm{t}_r\alpha_{q+r+1}$$

[**3**, p. 201; **4**, pp. 131–135].

In Elte's polytope $\Pi_6 = 1_{22}$ [**6**, Fig. 3.6a], the equatorial section $\Pi_5 = \mathrm{t}_1\beta_5$ has 10 facets $\beta_4$ and 32 facets $\mathrm{t}_1\,\alpha_4$ surrounding the two kinds of holes. Since the 32 $\mathrm{t}_1\,\alpha_4$'s are truncations of the facets of $\beta_5$, their centers are the vertices of a 5-cube $\gamma_5$, and 16 of them are just below the vertices of $\pi_5 = \mathrm{h}\gamma_5$, which is one of the cells of the higher $\mathrm{h}\delta_6$. We thus obtain the honeycomb $2_{22}$ and the lattice $E_6$.

The partition of 32 into $16 + 16$ provides nonlattice packings which have the same kissing number 72 and the same center density $1/8\sqrt{3}$ as $E_6$. Elte [**9**, p. 114] came close to discovering $2_{22}$ when he wrote: "It is impossible to bound a polytope by $\mathrm{V}_{27}$ only." (For him, unlike Kepler and Gosset [**10**], honeycombs were dismissed as would-be polytopes that fail to close up.)

Although the 54 facets $\mathrm{h}\gamma_5$ of $1_{22}$ are all congruent, we naturally name them alternately $1_{21}$ and $1_{12}$ so that there are 27 of each type. Similarly, the 6-dimensional honeycomb $2_{22}$ has congruent cells (Elte's $\mathrm{V}_{27}$) which may be named alternately $2_{21}$ and $2_{12}$. Each surrounds a hole in the lattice packing $E_6$. For the construction of a 7-dimensional packing $E_7$, we use a polytope $\Pi_7$ whose equatorial section $\Pi_6$ is $1_{22}$. The $27 + 27$ facets $1_{21}$ and $1_{12}$ of $1_{22}$ surround the two types of (congruent) holes. Those of the former type are just below the vertices of $\pi_6 = 2_{21}$, which is a cell of the higher $2_{22}$. Since the vertices of $\Pi_7$ are distributed in layers as $27 + 72 + 27$, their number is 126. In fact, this $\Pi_7$ is $2_{31}$ (Elte's $\mathrm{V}_{126}$ [**9**, p. 117; **6**, (3.75)]). The 7-dimensional honeycomb having vertex figure $2_{31}$ is $3_{31}$, and its vertices form the lattice $E_7$. Since there are two types of holes, the lattice packing is accompanied by nonlattice packings having the same kissing number 126 and center density $1/16$.

For the construction of an 8-dimensional packing $E_8$ based on layers $\Lambda_7 = E_7$, we use a polytope $\Pi_8$ whose equatorial section $\Pi_7$ is $2_{31}$. The 56 facets $2_{21}$ of $2_{31}$

surround 56 deep holes which are just below the vertices of $\pi_7 = 3_{21}$. This is a cell of the higher $3_{31}$. But now a new situation arises because $\Pi_8$ is Gosset's polytope $4_{21}$ [**10**; **6**, Fig. 3.8c] which has vertices not only on the "equator" and "tropics" but also at the "poles": $1 + 56 + 126 + 56 + 1$. Thus the cell $\pi_7 = 3_{21}$ of the higher $3_{31}$ is not a *facet* of $\Pi_8$ but its vertex figure! And $\Pi_8 = 4_{21}$ is the vertex figure of the honeycomb $5_{21}$ whose vertices form the lattice $E_8$. Like $D_4$, this packing is unique: there is no nonlattice packing which has the same kissing number 240 and center density $1/16$.

For the construction of a 9-dimensional packing $\Lambda_9$ based on layers $\Lambda_8 = E_8$, we use a polytope $\Pi_9$ whose equatorial section $\Pi_8$ is $4_{21}$. The 2160 facets $4_{11} = \beta_7$ of $4_{21}$ [**3**, p. 204] surround 2160 deep holes, only 16 of which can support balls in the next higher layer. So we use the distribution of the 2160 vertices of $2_{41}$ among 135 inscribed $\beta_8$'s. This indicates that the lattice packing $\Lambda_9$ is accompanied by non-lattice packings which have the same kissing number 272 and center density $1/16\sqrt{2}$. But now another new situation arises because $\Pi_9$ is not "uniform" but "axial": its symmetry group (only twice as big as that of $4_{21}$) is not transitive on its $16 + 240 + 16$ vertices. It is faintly reminiscent of the "partially truncated rhombohedron" in Dürer's famous *Melencolia* [**13**]. Moreover, another surprise, observed in 9 dimensions for the first time, is a nonlattice packing, called $P_{9a}$, in which some balls touch as many as 306 others (although the center density is less than $1/16\sqrt{2}$, namely, $5/128$; see page 140).

Still more remarkably, in 10 dimensions there is a non-lattice packing, called $P_{10c}$, whose kissing number 372 and center density $5/128$ exceed those of $\Lambda_{10}$. In 11 dimensions there is an arrangement of 582 balls all kissing one, although there is no systematic way to extend this local arrangement into a packing. In 12 dimensions the densest known packing, $K_{12}$, was not derived by lamination; it was discovered by J. A. Todd and the reviewer [**7**].

We must go up to 24 dimensions to see the most famous packing: the one provided by the *Leech lattice*, with kissing number 196560 and center density 1. Much of the book is inspired by the Leech lattice.

These remarks are intended to give some inkling of the rich subject of ball packing, although they only "scratch the surface" of the astonishing developments in the book. These include codes, hyperbolic geometry, Lie algebras, quadratic forms, geometry of numbers, Steiner systems, spherical $t$-designs, digital communication, and Monstrous Moonshine. The authors must be congratulated on their bibliographical research, which has produced more than fifteen hundred references!

REFERENCES

1. W. W. R. Ball and H. S. M. Coxeter, Mathematical Recreations and Essays, 13th ed., Dover, New York, 1987.
2. H. S. M. Coxeter, Extreme forms, *Canad. J. Math.*, 3 (1951) 391–441.
3. ______, Regular polytopes, 3rd ed., Dover, New York, 1973.
4. ______, Polytopes in the Netherlands, *Nieuw Archief voor Wiskunde*, (3) 26 (1978) 116–141.
5. ______, Regular and semi-regular polytopes. II, *Math. Z.*, 188 (1985) 559–591.
6. ______, Regular and semi-regular polytopes. III, *Math. Z.*, 200 (1989).
7. H. S. M. Coxeter and J. A. Todd, An extreme duodenary form, *Canad. J. Math.*, 5 (1953) 384–392.
8. C. Davis, B. Grünbaum, and F. A. Sherk, The Geometric Vein, Springer-Verlag, New York, Heidelberg, Berlin, 1981.
9. E. L. Elte, The Semiregular Polytopes of the Hyperspaces, Groningen, 1912.

10. T. Gosset, On the regular and semi-regular figures in space of $n$ dimensions, *Messenger of Math.*, 29 (1900) 43–48.
11. J. Kepler, The Six-cornered Snowflake, Clarendon Press, Oxford, 1966.
12. ______, L'Étrenne ou la Neige Sexangulaire, Librairie J. Vrin, Paris, 1975.
13. T. Lynch, The geometric body in Dürer's engraving *Melencolia I*.
14. M. Senechal and G. Fleck, Shaping Space: A Polyhedral Approach, Birkhäuser, Boston and Basel, 1988.
15. H. Steinhaus, Mathematical Snapshots, 2nd ed., Oxford University Press, 1950.

## *Geometric Inequalities.* By Yu. D. Burago and V. A. Zalgaller. Translated from the Russian by A. B. Sossinsky. Springer-Verlag, New York, 1988. xiv + 331 pp.

Don Chakerian
*Department of Mathematics, University of California, Davis, CA 95616*

**1. Historical sketch.** Geometric inequalities have long played a prominent role in the development of mathematics and science. As an early example, consider the ancient Egyptian approximation for the area of a quadrilateral, given by multiplying the averages of pairs of opposite sides. The reader may enjoy verifying that this is in fact an overestimate of the area, correct only when the quadrilateral is a rectangle (thus the pharaohs would have done well to encourage their surveyors to use this approximation when measuring farmland for purposes of taxation). The observation that the Egyptian formula is in excess of the exact area is only a step removed from a proof that a square encloses largest area among all quadrilaterals of the same perimeter. Here we have a precursor of the isoperimetric theorem, which states that a circle encloses maximum area among all simple closed plane curves of the same perimeter. This theorem, known to the early Greeks and the subject of a work by Zenodorus (circa 200 B.C.), has its equivalent expression in the most famous of geometric inequalities, the isoperimetric inequality,

$$L^2 \geqslant 4\pi A,$$

where $L$ is the perimeter of the given curve and $A$ the area enclosed.

In the early nineteenth century, Jacob Steiner carried the study of geometric inequalities to a new height, working in the purely synthetic tradition of Greek geometry. He introduced the powerful tool of Steiner symmetrization, which brings a figure closer to an optimal form by imposing more symmetry. Steiner used this in one of his proofs of the isoperimetric theorem, although he neglected to supply a proof that an optimizing figure exists. A bit later, H. A. Schwarz applied a different operation, Schwarz symmetrization, in a proof of the 3-dimensional version of the isoperimetric theorem, that a sphere encloses maximum volume among all simple closed surfaces of the same surface area. If the surface area is $S$ and the enclosed volume is $V$, this is equivalent to the inequality

$$S^3 \geqslant 36\pi V^2.$$

Hermann Minkowski, at the beginning of the present century, was instrumental in shaping a coherently structured theory of convex sets and geometric inequalities. While Minkowski's research was based on the work of earlier geometers, especially

that of Hermann Brunn, he went far beyond them in rigor and organization. One of his greatest contributions was the discovery of mixed volumes, vastly generalizing the concepts of volume, surface area, and total mean curvature (his initial ideas on the subject are described in a letter [**5**, pp. 133-136] to Hilbert dated Dec. 10, 1900). The mixed volume $V(K_1,\dots,K_n)$ of convex bodies $K_1,\dots,K_n$ in Euclidean $n$-space $R^n$ is invariant under translations of the $K_i$ and multilinear with respect to the operations of scalar multiplication and vector addition of sets. The definition is such that $V(K,\dots,K)$ is the volume of $K$, and $nV(K,\dots,K,B)$ is the surface area of $K$ when $B$ is the unit $n$-ball; thus the isoperimetric theorem and its higher dimensional counterparts all turn out to be very special cases of inequalities between mixed volumes first proved by Minkowski.

The elegant 1916 monograph by Wilhelm Blaschke [**1**] contributed to the synthesis of the work of Steiner, Schwarz, Brunn, and Minkowski and gave a sense of the classic beauty of the subject. The progress in convexity and geometric inequalities up to 1934 was lucidly summarized in the book of Bonnesen and Fenchel [**2**]. That work, an example of sterling exposition of a technical subject, has only recently been translated into English.

In 1936, A. D. Alexandrov and, independently, Werner Fenchel, established an important extension of the Minkowski inequalities for mixed volumes. The Alexandrov-Fenchel inequality states, for convex bodies $K_1,\dots,K_n$ in $R^n$, that

$$V(K_1, K_2, K_3,\dots, K_n)^2 \geqslant V(K_1, K_1, K_3,\dots, K_n)V(K_2, K_2, K_3,\dots, K_n).$$

It is interesting that the conditions necessary for equality are still not completely settled. In case $K$ is a plane convex set of perimeter $L$ and area $A$, we have $V(K, K) = A$ and $2V(K, B) = L$, where $B$ is the planar unit disk. Thus the Alexandrov-Fenchel inequality, with $n = 2$, $K_1 = K$ and $K_2 = B$, becomes the isoperimetric inequality.

In connection with the more recent history of the field, the work of Hugo Hadwiger has had a strong unifying effect. His little book [**4**] is another expository gem and a worthy companion to Blaschke's *Kreis und Kugel*. However, to acquire a sense of how art, form and symmetry combine to give the subject its rich beauty, one can do no better than read the magnificent book of L. Fejes Tóth, *Regular Figures*, [**3**].

**2. The book under review.** Burago and Zalgaller concern themselves with what has been accomplished since the appearance of Bonnesen and Fenchel's 1934 classic. Their book, while accessible to a broad range of mathematicians, will be of greatest interest to specialists in geometric inequalities, providing an invaluable reference for recent developments in the field. We should point out that the book emphasizes the differential-geometric aspects of the subject rather than problems of a more combinatorial flavor, as exemplified, for instance, in Fejes Tóth's *Regular Figures*.

After a quick overview of the multitude of proofs that have been given for the isoperimetric theorem in the plane, the authors proceed to a detailed treatment of the relationship between isoperimetric inequalities and curvature on surfaces. The extensions of these ideas to Riemannian manifolds are fully developed in the latter part of the book, including problems such as estimating the volume of a manifold with given constraints on the curvature. A careful statement of the higher dimen-

sional version of the isoperimetric theorem in fullest generality demands a precise definition of surface area. The authors provide an excellent survey of the required notions from geometric measure theory in the course of their discussion of the modern descendents of the isoperimetric theorem.

The chapter on mixed volumes, written in collaboration with V.P. Fedotov, gives a thorough introduction to the theory of mixed volumes and the latest extensions of the concept. An addendum to this chapter, contributed by A. G. Khovanskiĭ, is devoted to a lucidly presented proof of the Alexandrov-Fenchel inequality based on machinery from algebraic geometry, involving in particular the Hodge inequality for the intersection index of a pair of curves. This is among the very surprising recent discoveries of connections between mixed volumes and other seemingly far removed parts of mathematics. Another surprise is the appearance of ideas related to mixed volumes in the recent proof of van der Waerden's conjecture concerning the minimum value of the permanent of a stochastic matrix. The original ideas of Minkowski are presently finding growing application in a variety of areas, including probability theory and functional analysis.

**3. Geometric inequalities in the curriculum.** The authors remark, on page 165, that "Mixed volumes are not included in traditional teaching syllabuses and the literature on this topic is not very diversified."

One may add that in the United States the general topic of geometric inequalities goes unmentioned in the curriculum at all levels. This despite the fact that the isoperimetric inequality and its relatives can be treated from an elementary viewpoint accessible even to secondary school students, while still conveying the power of the subject in its diversity of application and simultaneously displaying its intrinsic charm and beauty. If high schoolers were given a taste of such fare, as a substitute for some of the rather repellent portions of the menu we now place before them, they might enter our college classrooms with whetted appetites, demanding more of this wonderful mathematical tradition. There then might be a revival of interest in geometry, and we might yet be able to rescue from oblivion that precious heritage given to us by the Greeks, the art of visual thinking.

REFERENCES

1. W. Blaschke, Kreis und Kugel, 1916, second edition, de Gruyter, Berlin, 1956.
2. T. Bonnesen and W. Fenchel, Theorie der konvexen Körper, Springer, Berlin, 1934. English translation by L. Boron, C. Christenson and B. Smith, in collaboration with W. Fenchel, BCS Associates, Moscow, Idaho, 1987.
3. L. Fejes Tóth, Regular Figures, Pergamon, Oxford, 1964.
4. H. Hadwiger, Altes und Neues über konvexe Körper, Birkhäuser, Basel and Stuttgart, 1955.
5. H. Minkowski, Briefe an Hilbert, Springer, Berlin, 1973.

## TELEGRAPHIC REVIEWS

*Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.*

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

| | | |
|---|---|---|
| T: Textbook | P: Professional Reading | 1–4: Semesters |
| C: Computer Software | L: Undergraduate Library | **: Special Emphasis |
| S: Supplementary Reading | 13: Grade Level | ??: Questionable |

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

**General, L**.** *U.S.A. Mathematical Olympiads, 1972-1986.* Murray S. Klamkin. New Math. Lib., V. 33. MAA, 1988, xv + 127 pp, $13.50 (P). [ISBN: 0-88385-634-4] Begun in 1972, the Olympiad is the high school equivalent of the undergraduate Putnam contest. Following the fifteen contests are solutions, along with further comments, grouped by subject. Appendices contain a list of winners, a list of commonly used definitions and theorems, and an extensive bibliography valuable to problem solvers. GG

**General, S**(13).** *Mathematical Formulas: Algebra, Geometry, Mathematical Analysis.* A.G. Tsypkin, G.G. Tsypkin. MIR (US Distr: Imported Pub), 1988, 183 pp, $4.95 (P). [ISBN: 5-03-000806-3] Excellent collection of formulas from algebra, geometry (analytic, differential, and vector), theory of functions, etc. Includes a table of over 400 integrals. At $4.95 and pocket-sized, it makes a handy reference book for undergraduates. MR

**General, P, L.** *Collected Mathematical Papers.* Igor R. Shafarevich. Springer-Verlag, 1989, vi + 769 pp, $99. [ISBN: 0-387-13618-5] Contains the authors papers which were published between 1943 and 1984. Of the 43 papers, 41 are in English translation and one each in German and French. CEC

**General, L.** *Thomas Gray, Philosopher Cat.* Philip J. Davis. Harcourt Brace Jovanovich, 1988, xi + 143 pp, $10.95. [ISBN: 0-15-188100-6] The delightful tale of a slightly eccentric Fellow of Pembroke College and a cat with a philosophical bent. Reveals the scholarly life and attitudes at Cambridge in a story wrapped lightly around a mathematical problem. CEC

**General, P.** *Eleven Papers Translated from the Russian.* V.I. Bernik, *et al.* Transl. Ser. 2, V. 140. AMS, 1988, viii + 147 pp, $50. [ISBN: 0-8218-3116-X] Papers on logic, algebra, number theory, and optimisation; two are forty years old, others date from 1982-1985. LAS

**General, L.** *Gardner's Whys & Wherefores.* Martin Gardner. U of Chicago Pr, 1989, ix + 261 pp, $19.95. [ISBN: 0-226-28245-7] A collection of sixteen essays and twenty book reviews, most written within the last six years, covering an enormous range of topics: annotations to Casey at the Bat, essays about pi and the travelling salesman from *Discover*, reviews of Gleik's *Chaos*, Feynman's autobiography, and Bloom's *Closing of the American Mind.* Each reflects Gardner's passion for interesting detail and wry reflections on logical twists. One highlight is a devastating review by Gardner of one of his own books, which appeared under a pseudonym in the *New York Review of Books.* LAS

**Elementary, T(13: 1).** *Plane Trigonometry, Fifth Edition.* Bernard J. Rice, Jerry D. Strange. PWS-Kent, 1989, x + 358 pp. [ISBN: 0-534-91562-0] With this edition, the use of mathematical tables may be omitted if a "pure calculator" course is desired. Otherwise, this edition is essentially the same as the previous edition. (*First Edition*, TR, August-September 1975; *Second Edition*, TR, October 1978; *Fourth Edition*, TR, April 1987.) LCL

**Mathematics Appreciation, T(13: 1).** *A Survey of Mathematics with Applications, Third Edition.* Allen R. Angel, Stuart R. Porter. Addison-Wesley, 1989, xii + 819 pp, $37.50. [ISBN: 0-201-13696-1] Changes in this edition include the addition of a second chapter on algebra and a new chapter on problem solving, new exercises on problem solving and research questions, and new chapter tests. (*Second Edition*, TR, April 1986.) JNC

**Precalculus, S(13), L*.** *Elementary Mathematics: Selected Topics and Problem Solving.* G. Dorofeev, M. Potapov, N. Rosov. MIR (US Distr: Imported Pub), 1988, 488 pp, $15.95. Valuable reference for teachers of algebra, trigonometry, and geometry. Textual material illustrates methods of solving problems and points out common student errors. Extensive exercises, with answers, pitched slightly higher than typical textbook. GG

**Precalculus, T(13).** *Fundamentals of College Algebra, Seventh Edition.* Earl W. Swokowski. PWS-Kent, 1989, x + 463 pp. [ISBN: 0-534-91748-8] An old standard enters its seventh edition with yet another permutation of the binomial theorem, graphing, inverses, logarithms, sequences, etc., and with an ever-expanding family of supplements that include tests on disks, transparencies, Kemeny's precalculus software. (*Third Edition,* TR, January 1976; *Fourth Edition,* TR, January 1979.) AWR

**Precalculus, T**(13), S*.** *College Trigonometry.* Stanley I. Grossman. Saunders College, 1989, xv + 353 pp [ISBN: 0-03-007103-8]; *Algebra and Trigonometry,* 1989, xvii + 726 pp. [ISBN: 0-03-007129-1] Includes many examples with every step included, and applications. Numerous exercises, many calculator-oriented, including chapter review. Answers provided to most odd-numbered exercises. Includes index of applications in the sciences, business, and economics. Many sections contain biographical and historical anecdotes. *Algebra and Trigonometry* covers usual topics in algebra, functions and graphs, exponential and logarithmic functions, and trigonometry. Also includes introductions to linear algebra and discrete mathematics. *College Trigonometry* includes a nice algebra review, complete discussion of angles, trigonometric functions and graphs, identities and applications, complex numbers, exponential and logarithmic functions. Good discussion of inverse trigonometric functions. SB

**Precalculus, T(13: 1).** *College Algebra and Trigonometry.* Louis Leithold. Addison-Wesley, 1989, xviii + 781 pp, $40.75. [ISBN: 0-201-15730-6] Comprehensive text that forms the backbone of a new modular series of precalculus books. The standard topics for such a course are here: algebraic review, polynomial, exponential, logarithmic, trigonometric functions, and their graphs, as well as treatments of vectors, polar coordinates, systems of linear equations/inequalities, and matrices. A final chapter covers induction, sequences, series, combinatorial probability, and the binomial theorem. Numerous worked examples and section exercises. Answers to odd-numbered exercises. CE

**Precalculus, T(13: 1).** *College Algebra.* Philip Gillett. Scott Foresman, 1989, 484 pp, $28.50 [ISBN: 0-673-18793-4]; *Algebra and Trigonometry,* 1989, 664 pp, $29.50. [ISBN: 0-673-18311-4] A well-written, traditional approach with interesting historical references and interspersed, marginal, true-false questions. The algebra content of the two texts is identical—the extra dollar buys three chapters on trigonometry. JNC

**Precalculus, T(13: 1).** *Algebra and Trigonometry.* Marvin L. Bittinger, Judith A. Beecher. Addison-Wesley, 1989, xvii + 731 pp, $37.50 (P). [ISBN: 0-201-09152-6] A traditional approach emphasizing real-life applications and graphing. Includes calculator exercises and interspersed margin exercises. Available supplements include fifty video tapes of text material and printed and computerized test banks. JNC

**Precalculus, T(13: 1).** *Trigonometry, Fourth Edition.* Margaret L. Lial, Charles D. Miller. Scott Foresman, 1989, 406 pp, $24.95. [ISBN: 0-673-38248-6] Clean and lean. Early review of needed geometry. Trigonometric functions of general angles and identities are introduced early on. Some trigonometry of acute- and right-angled triangles follows. Then come radian measure, graphing, and so on, through vectors, complex numbers, polar coordinates, and the exponential and logarithmic functions. Brief tables. Some calculator exercises. Inside front and back covers have helpful summaries of main results. Pace is leisurely and the writing is clear. (*First Edition,* TR, November 1977; *Second Edition,* TR, August-September 1981; *Third Edition,* TR, November 1985.) JK

**Precalculus, T(13: 1).** *Precalculus.* Joseph Elich, Lawrence O. Cannon. Scott Foresman, 1989, 670 pp, $29. [ISBN: 0-673-18831-0] Introduces number systems, functions, inverses, exponentials, logarithms, trigonometric functions, systems, analytic geometry, and discrete mathematics. Many helpful two-color figures. Exercise sets have drill problems, application (word) problems, and a set of more open-ended exploration problems. Includes 31 historical sketches and an appendix on using a calculator. SP

**Finite Mathematics, T(2), S.** *Finite Mathematics and Calculus with Applications, Third Edition.* Margaret L. Lial, Charles D. Miller. Scott Foresman, 1989, xxii + 1098 pp, $35. [ISBN: 0-673-38255-9] This edition places greater emphasis on business and economics, and includes new examples, exercises, and applications. Routine exercises are now followed by application exercises arranged by discipline. Previous editions were published under the title *Mathematics and Calculus with Applications* (*First Edition,* TR, December 1980; *Second Edition,* TR, March 1986). JNC

**Finite Mathematics, T(13-14: 1).** *Methods of Finite Mathematics.* John W. Brown, Donald R. Sherbert. Wiley, 1989, xii + 668 pp, $48.16. [ISBN: 0-471-63003-9] Linear methods in the plane, systems of linear equations, matrices, linear programming, methods of counting, methods of probability, probability distributions and statistics, Markov analysis, mathematics of finance, game theory, logic and graphs. If you teach a course covering a combination of these topics you might consider this book. Numerous examples and exercises, along with the usual pedagogical aids. Software package with manual is available. Medium level difficulty. JK

**Finite Mathematics, T(13).** *Finite Mathematics, Fourth Edition.* Margaret L. Lial, Charles D. Miller. Scott Foresman, 1989, xvii + 572 pp, $33. [ISBN: 0-673-38253-2] Here is another mix of matrices, linear programming, Markov chains, sets, counting, probability and statistics, and a chapter on finance. More exercises, more applications; has edition $n+1$ ever had less of these than edition $n$? (*Second

*Edition*, TR, August-September 1982; *Third Edition*, TR, January 1986.) AWR

**Education, S, L.** *Board Games Round the World: A Resource Book for Mathematical Investigations.* Robbie Bell, Michael Cornelius. Cambridge U Pr, 1988, 124 pp, $9.95 (P). [ISBN: 0-521-35924-4] Brief descriptions of several dozen board games, grouped by type (position, Mancala, war, race, dice) with suggestions for investigations of strategy and features ("specialise, generalise, conjecture, convince") that will help school children learn to think mathematically. LAS

**Education, P, L.** *Mathematics: Report of the Project 2061 Phase I Mathematics Panel.* David Blackwell, Leon Henkin. AAAS, 1989, xi + 47 pp, $7.50 (P). [ISBN: 0-87168-344-X] One of five panel reports written as background for the AAAS Project 2061, this report offers an ambitious vision of which important ideas of mathematics should be understood by next century's "typical adults"—those with no mathematics education beyond high school. In addition to the expected elementary topics from the present school curriculum, this panel's view of mathematical literacy for all Americans includes Turing machines, social choice functions, simple systems of differential equations, and linear programming. The example of Garfield High School is cited in support of a discussion of emotions and mathematics to argue that such a vision is not unrealistic. Nevertheless, the full report *Science for All Americans* (see below) includes relatively little of the more innovative recommendations in this report. LAS

**Education, P, L*.** *Science For All Americans.* AAAS, 1989, ix + 217 pp, $14.50 (P). [ISBN: 0-87168-341-5] Summary report of Project 2061, the futuristic attempt by AAAS to articulate what every person should know of mathematics, science, and technology. Twelve core chapters summarise the recommendations in mini-essays covering the entire spectrum of science. Mathematics appears throughout, both linked with science and in three special sections: on mathematical ways of thinking (abstraction, manipulation, application); on the mathematical world (numbers, symbols, shapes, uncertainty, data, sampling, reasoning); and on mathematical skills (computation, calculator, estimation). The vision of mathematics for all Americans contained in this summary is substantially stronger than current average performance, but much weaker than that recommended by the project's own mathematics panel (see above), or by the new NCTM *Standards* for school mathematics (see below). LAS

**Education, P**, L**.** *Curriculum and Evaluation Standards for School Mathematics.* NCTM, 1989, vii + 258 pp, $25 (P). [ISBN: 0-87353-273-2] "Must" reading for anyone who teaches mathematics. Presents guidelines for the content of school mathematics curricula and suggests means for evaluating student mastery of this content. Although the K–12 curriculum is the primary focus, the goals articulated in the *Standards* can (and should!) inform the undergraduate mathematics curriculum as well. AO

**Education, S*, P*, L***.** *Escalante: The Best Teacher in America.* Jay Mathews. Henry Holt, 1988, ix + 322 pp, $19.95. [ISBN: 0-8050-0450-5] A fascinating, behind-the-scenes portrait of Jaime Escalante, his students, and fellow teachers at Garfield High School in East Los Angeles—a school which moved in ten years from intellectual bankruptcy to outstanding success in AP calculus and other AP courses. At times a hard-driving coach, cult hero, and stubborn individualist, Escalante found and reinforced *ganas* ("the urge") in the most unlikely students. An eccentric genius, Escalante's idiosyncratic methods ("Escalante made calculus into a religion") reveal more about what is possible than about what is practical. Nevertheless, this story, like the movie *Stand and Deliver*, is an impressive celebration of the art of teaching. LAS

**Education, P, L.** *American Perspectives on the Sixth International Congress on Mathematical Education.* Ed: Thomas J. Cooney. NCTM, 1989, v + 56 pp, $6.75 (P). [ISBN: 0-87353-276-7] A mélange of commentary by twenty-four American participants in ICMI-VI held in Budapest from July 27 to August 3, 1988. Six authors—J. Kilpatrick, J. Fey, L. Steffe, N. Kreinberg, R. Shumway, and P. Zorn—offer broad surveys; other shorter papers provide editorial observations based on personal experiences with the vast Congress. Topics include technology, teacher education, visualisation, contests, community colleges, and college mathematics. LAS

**History, L.** *The First Electronic Computer: The Atanasoff Story.* Alice R. Burks, Arthur W. Burks. U of Michigan Pr, 1988, xii + 387 pp, $30. [ISBN: 0-472-10090-4] Covers the story of the electronic computer completed by John Atanasoff in 1942. Based on the authors' direct knowledge and on the proceedings of the Honeywell vs. Sperry Rand case, which concluded that Atanasoff's computer was an automatic electronic digital computer prior to the ENIAC computer which was unveiled in 1946, and that the ENIAC had been derived from it. RH

**Logic, P.** *Logic Colloquium '84.* Ed: J.B. Paris, A.J. Wilkie, G.M. Wilmers. Stud. in Logic & the Found. of Math., V. 120. North-Holland (US Distr: Elsevier Science), 1986, x + 377 pp, $66.75. [ISBN: 0-444-87999-4] Research and survey papers presented at the 1984 European Summer Meeting of the Association for Symbolic Logic, held at the University of Manchester, July 15-24. Main topics are model theory of arithmetic and algebra, semantics of natural languages, and applications of logic to complexity theory. KS

**Foundations, P, L.** *The Liar: An Essay on Truth and Circularity.* Jon Barwise, John Etchemendy. Oxford U Pr, 1987, xii + 185 pp. [ISBN: 0-19-505072-X] A fresh discussion of the ancient liar paradox ("This sentence is false.") from two related perspectives: the context-sensitive logic introduced in 1950 by John Austin in which the paradox is re-

solved, in analogy with Russell's hierarchy of sets, by making truth depend on levels of context; and the new theory of non-well-founded sets introduced in 1987 by Peter Acsél in which an "anti-foundation axiom" (AFA) based on "decorating" graphs permits a consistent theory of hypersets in which there is a unique set $\Omega$ equal to its own singleton ($\Omega = \{\Omega\}$). Opens many avenues of further investigation into foundations of linguistics, computer science, set theory, and philosophy. LAS

**Graph Theory, P.** *Proceedings of the First Japan Conference on Graph Theory and Applications.* J. Akiyama, Y. Egawa, H. Enomoto. Annals of Disc. Math., V. 38. North-Holland (US Distr: Elsevier Science), 1988, vi + 418 pp, $122. [ISBN: 0-444-70538-4] A collection of 46 papers printed in a very attractive format. AWR

**Graph Theory, P.** *Applications of Graphs in Chemistry and Physics.* Ed: John W. Kennedy, Louis V. Quintas. North-Holland (US Distr: Elsevier Science), 1988, 416 pp, $123.75. [ISBN: 0-444-70513-9] Graph theory—the combinatorics and topology of nodes and edges—permeates and percolates through physical chemistry. The twenty-five papers here are largely expository, but not for the graph-theory novice. Note the price! BC

**Combinatorics, T*(15-18: 1), S*, P, L*.** *Combinatorial Search.* Martin Aigner. Ser. in Comput. Sci. Wiley, 1988, 368 pp, $44.95. [ISBN: 0-471-92142-4] The purpose of this book is to give an introduction to the basic ideas (Chapter 1), and present a collection of the most interesting search problems (Chapters 2-6, independently arranged: Weighing Problems, Graph Problems, Sorting Problems, Poset Problems, More Problems). Each section is followed with a set of exercises and solutions to the most interesting, and each chapter concludes with a list of unsolved problems and a guide to further reading. LCL

**Discrete Mathematics, T(14-15: 1, 2).** *Discrete and Combinatorial Mathematics: An Applied Introduction, Second Edition.* Ralph P. Grimaldi. Addison-Wesley, 1989, xvi + 815 pp, $43. [ISBN: 0-201-11954-4] Emphasis on algorithms and applications; most theorems proved. Covers counting, logic, set theory, integers and induction, relations and functions, finite state machines, generating functions, recurrence relations, graph theory, trees, optimisation, modular arithmetic, Boolean algebra, coding theory, finite fields, Pólya's method of enumeration. Each chapter ends with summary and historical review. Many examples and exercises (solutions to odd-numbered exercises). Includes computer science applications. SB

**Number Theory, P*, L.** *Ramanujan's Notebooks, Part II.* Bruce C. Berndt. Springer-Verlag, 1989, xi + 359 pp, $79.80. [ISBN: 0-387-96794-X] The second of four volumes devoted to the editing of the *Notebooks.* Chapters 10-15 of the second notebook are examined. Except in a few instances, each result in these chapters is established. CEC

**Number Theory, P.** *Conjecture de Lehmer et Petits Nombres de Salem.* Marie José Bertin, Martine Pathiaux-Delefosse. Papers in Pure & Appl. Math., No. 81. Queen's U, 1989, 144 pp, (P). An exposition of some recent results on a question of Lehmer concerning the roots of certain polynomials with integer coefficients and the connection between this question and so-called Salem numbers and Pisot numbers. SG

**Number Theory, P.** *Algebraic K-Theory and Algebraic Number Theory.* Ed: Michael R. Stein, R. Keith Dennis. Contemp. Math., V. 83. AMS, 1989, xiv + 488 pp, $46 (P). [ISBN: 0-8218-5090-3] Nineteen papers originating from joint U.S.-Japan seminar held in Honolulu in January 1987. GG

**Number Theory, T(16-17: 1).** *Introduction to Number Theory.* Daniel E. Flath. Wiley, 1989, xii + 212 pp, $39.95. [ISBN: 0-471-60836-X] Topics include prime numbers and unique factorisation, sums of two squares, quadratic reciprocity, indefinite forms, and the class group and genera. Emphasis is on Diophantine equations, especially quadratic equations in two variables. Includes numerous exercises and an appendix in which Jean-Pierre Serre discusses related current research. RH

**Number Theory, P.** *Representation Theory and Number Theory in Connection with the Local Langlands Conjecture.* Ed: J. Ritter. Contemp. Math., V. 86. AMS, 1989, xiii + 266 pp, $28 (P). [ISBN: 0-8218-5093-8] Proceedings of a conference held at the University of Augsburg on December 8-14, 1985. Includes 21 papers. RH

**Number Theory, P.** *Topological Methods in Galois Representation Theory.* Victor P. Snaith. Canadian Math. Soc. Ser. of Mono. & Adv. Texts. Wiley, 1989, xiii + 299 pp, $44.95. [ISBN: 0-471-61752-0] An up-to-date exposition of the study of topological invariants of representations that arise in the Galois theory of fields. The author begins with a brief description of abelian and nonabelian cohomology of groups. After presenting characteristic classes, he presents a form of Brauer's induction theorem and some of its applications to number theory. SG

**Linear Algebra, T(14: 1).** *Elementary Linear Algebra, Third Edition.* Stewart Venit, Wayne Bishop. PWS-Kent, 1989, xiv + 462 pp. [ISBN: 0-534-91689-9] Printed in a new, attractive two-color format; the only substantial change is the addition of interspersed "mini-applications." (TR, *First Edition,* April 1981; TR, *Second Edition,* June-July 1987.) JNC

**Linear Algebra, T*(14: 1).** *Introduction to Linear Algebra, Second Edition.* Lee W. Johnson, R. Dean Riess, Jimmy T. Arnold. Addison-Wesley, 1989, xi + 578 pp, $40.75. [ISBN: 0-201-16833-2] Gradually increases the level of abstraction; begins with matrix theory and systems of linear equations; introduces vector-space ideas in $R^n$ before generalised vector spaces: includes comments on computer solutions and a chapter on numerical methods. This edition has more explanations, motivation, examples, and exercises and covers eigenvalues in two chap-

ters. JNC

**Linear Algebra, T(14: 1).** *Linear Algebra with Applications, Third Edition.* Jeanne L. Agnew, Robert C. Knapp. Brooks/Cole, 1989, viii + 392 pp, $36.50. [ISBN: 0-534-09456-2] Sophomore-level text whose early introduction of eigenvalues (before vector spaces and linear transformations) distinguishes it from most. Plentiful exercises are mostly routine, with answers to odd problems provided. Main content change from first two editions (TR, *First Edition,* January 1979; TR, *Second Edition,* October 1983; Extended Review, November 1986) is deletion of chapter on linear programming. GG

**Linear Algebra, S(14), L.** *Schaum's Outline of Theory and Problems of Matrix Operators.* Richard Bronson. McGraw-Hill, 1989, 230 pp, $8.95 (P). [ISBN: 0-07-007978-1] Features an algorithmic approach with emphasis on computationally efficient methods. JNC

**Linear Algebra, S(13).** *Systems of Linear Equations.* L.A. Skornyakov. Little Math. Lib. MIR (US Distr: Imported Pub), 1988, 64 pp, $1.95 (P). [ISBN: 15-03-000268-5] Bases the theory of systems of linear equations on elementary operations on matrices. Translated and revised from the 1986 Russian edition. Exercises. JNC

**Linear Algebra, S(13-14).** *Primer for Linear Algebra.* Stephen Demko. Scott Foresman, 1989, 189 pp, $8 (P). [ISBN: 0-673-38642-2] A bare-bones treatment of linear systems, matrices, linear independence, determinants, and eigenvalues. Written in part with multivariable calculus in mind, contains sections on the second derivative test, Jacobians, Newton's method for nonlinear systems, tangent planes, and fractals. Surprisingly wide range, in a computational sense. No abstraction or proofs. GG

**Linear Algebra, S*.** *Schaum's Solved Problems Series: 3000 Solved Problems in Linear Algebra.* Seymour Lipschutz. McGraw-Hill, 1989, vii + 480 pp, $19.95 (P). [ISBN: 0-07-038023-6] This collection covers almost every type of linear algebra problem, both computational and theoretical in nature. Useful as a supplement or as a refresher. CEC

**Group Theory, P*, L**.** *Elements of Mathematics: Lie Groups and Lie Algebras, Chapters 1-3.* Nicolas Bourbaki. Springer-Verlag, 1989, xvii + 450 pp, $59. [ISBN: 0-387-50218-1] English translation of the first part of Bourbaki's *Groupes et algèbres de Lie,* covering Lie algebras, free Lie algebras, and Lie groups. Valuable, historical notes; voluminous exercises. (TR of 1971 French edition, May 1972.) LAS

**Algebra, P.** *Lecture Notes in Mathematics-1352: Algebra, Some Current Trends.* Ed: L.L. Avramov, K.B. Tchakerian. Springer-Verlag, 1988, ix + 239 pp, $20 (P). [ISBN: 0-387-50371-4] Proceedings of the Fifth National School in Algebra held on the Black Sea coast near Varna, Bulgaria, from September 24 to October 4, 1986. LAS

**Algebra, T(13: 1).** *College Algebra.* Stanley I. Grossman. Saunders College, 1989, xvii + 503 pp. [ISBN: 0-03-007089-9] A textbook designed to hold the interest of both recent high school graduates and students who have been away from institutionalised education for some time. The book contains approximately 125 "real world" applications, and numerous amusing vignettes from the history of mathematics. Many examples (600+) and exercises (3500+). Pleasant and unintimidating design. SM

**Algebra, P*, L**.** *Elements of Mathematics: Algebra I, Chapters 1-3.* Nicolas Bourbaki. Springer-Verlag, 1989, xxiii + 708 pp, $69. [ISBN: 0-387-19373-1] Reprint of Addison-Wesley's 1974 translation (TR, October 1974) of the first three chapters of Bourbaki's *Algèbre* originally written two decades earlier. This volume treats algebraic structures, linear algebra, and tensor algebras. Challenging exercises, important historical notes. A worthwhile volume to make available to libraries and young researchers. LAS

**Algebra, P*, L**.** *Elements of Mathematics: Commutative Algebra, Chapters 1-7.* Nicolas Bourbaki. Springer-Verlag, 1989, xxiv + 625 pp, $69. [ISBN: 0-387-19371-5] Reprint of Addison-Wesley's 1972 translation (TR, June 1973) of *Éléments de mathématique, Algèbre commutative.* Seven chapters beginning with flat modules, concluding with divisors; concludes with a rich, historical essay and tables of implications and invariances. LAS

**Algebra, P.** *Ordered Sets and Lattices.* Kh. Drashkovicheva, *et al.* AMS Transl. Ser. 2, V. 141. AMS, 1989, x + 203 pp, $69. [ISBN: 0-8218-3121-6] A translation of five articles originally published in Russian describe (nearly all) the results in lattice theory in the past fifteen years—"reviews of reviews" according to the translator. If you want to know what's happening in lattice theory, start here. SG

**Algebra, P.** *Classical Groups and Related Topics.* Ed: Alexander J. Hahn, Donald G. James, Zhe-xian Wan. Contemp. Math., V. 82. AMS, 1989, xvi + 254 pp, $28 (P). [ISBN: 0-8218-5089-x] Contains several interesting papers on many aspects of classical groups that were delivered at a 1987 conference in honor of the memory of L.K. Hua. SG

**Algebra, P.** *Unit Groups of Classical Rings.* Gregory Karpilovsky. Clarendon Pr, 1988, xiv + 370 pp, $98. [ISBN: 0-19-853557-0] Account of what is known of the structure of the unit group $U(R)$ of a ring $R$, including many recent results. Motivated by problems of determining the isomorphism class of $U(R)$ in terms of natural $R$ invariants and finding an effective method for constructing units of $R$. Covers algebraic units, unit groups of cyclotomic fields, multiplicative groups of local and global fields and extensions, multiplicative groups of division rings, cyclic unit groups, finite generation of $U(R)$, unit groups of group rings. SB

**Calculus, T(13-14: 2, 3).** *Calculus with Analytic Geometry, Fourth Edition.* Murray H. Protter, Philip E. Protter. Jones & Bartlett, 1988, xii + 962 pp, $50. [ISBN: 0-86720-093-6] Changes in

this edition (*Second Edition*, TR, August-September 1970; *Third Edition*, TR, October 1977) include the addition of two new chapters: Chapter 17 on vector field theory, and Chapter 18 on differential equations. Other changes include a thorough review of precalculus material in Chapter 1, a reordering of some topics, expanded coverage of infinite series, and new sections on Newton's method and on applications of the integral to probability problems. More emphasis given to the graphing of functions as surfaces. The number of problems has been increased by more than one-third, and review problems have been added to the end of each chapter. RH

**Calculus, S(13).** ***Misteaks ... and how to find them before the teacher does ..., Second Edition.*** Barry Cipra. Academic Pr, 1989, xiii + 65 pp, $6.95 (P). [ISBN: 0-12-174695-X] This is a cute little book that can be summarised as an effort to get students to ask some "obvious" questions about the reasonableness of their answers. How do we get students to read it? (*First Edition*, TR, March 1984.) AWR

**Calculus, T(13: 1).** ***Calculus with Applications, Fourth Edition.*** Margaret L. Lial, Charles D. Miller. Scott Foresman, 1989, xviii + 794 pp, $34. [ISBN: 0-673-38251-6] For majors in business, management, economics, life, or social sciences. New features include motivational material, additional examples, exercises, and applications (particularly from business and economics). Routine exercises are now followed by clearly marked application exercises arranged by discipline. (*Third Edition*, TR, March 1986). JNC

**Calculus, T(13-14: 2).** ***Mathematics and Calculus with Applications.*** Marvin L. Bittinger, J. Conrad Crown. Addison-Wesley, 1989, xxv + 1115 pp, $44. [ISBN: 0-201-05941-X] An introductory text combining finite mathematics and business calculus. Contains a variety of exercises including extended applications, calculator, and computer exercises. Introduction to linear systems, linear programming, probability, as well as differential and integral calculus. SP

**Calculus, T(13).** ***Mathematical Analysis.*** V.B. Uvarov. MIR (US Distr: Imported Pub), 1988, 335 pp, $9.95. [ISBN: 5-03-000500-5] Here is the ultimate lean and lively calculus text. A little more rigorous than American texts in its treatment of the real and complex numbers, this book marches through all the usual topics of a three-term calculus sequence, covers Jordan-measurable sets in its approach to multiple integrals, and throws in a chapter on Fourier series. This is accomplished in 335 small pages, and is available for under $10 in hardcover! AWR

**Real Analysis, S(17-18), L*.** ***The Fourier Integral and Certain of its Applications.*** Norbert Wiener. Cambridge U Pr, 1988, xvii + 201 pp, $16.95 (P). [ISBN: 0-521-35884-1] Third printing of a true classic, described in a new Foreword by Jean-Pierre Kahane as "one of the great mathematical achievements of this century." Wiener gives a unique personal view centered on Tauberian theorems which served to introduce the idea of Banach algebras and abstract harmonic analysis decades before the modern theory emerged. First published in 1933, based on a series of lectures given at Cambridge University. LAS

**Complex Analysis, P.** ***Holomorphic Functions and Moduli I & II.*** Ed: D. Drasin, *et al.* Springer-Verlag, 1988. *Moduli I*, Math. Sci. Res. Inst. Publicat., V. 10, xiii + 246 pp, $29.50 [ISBN: 0-387-96766-4]; *Moduli II*, Math. Sci. Res. Inst. Publicat., V. 11, xiii + 290 pp, $32.80. [ISBN: 0-387-96786-9] Collection of papers solicited from participants in the Spring 1986 Program in Geometric Function Theory at the Mathematical Sciences Research Institute in Berkeley, California. Topics include Teichmüller theory, quasiconformal mappings, Kleinian groups, univalent functions, and value distributions. BH

**Differential Equations, T(18: 1), S, P.** ***Critical Point Theory and Hamiltonian Systems.*** Jean Mawhin, Michel Willem. Appl. Math. Sci., V. 74. Springer-Verlag, 1989, xiv + 277 pp, $54. [ISBN: 0-387-96908-X] After a brief historical sketch, this text develops the fundamentals of critical point theory, and eventually focuses somewhat on periodic solutions problems for Hamiltonian systems. Written for those trained in ordinary differential equations. MU

**Differential Equations, T(15-16: 1).** ***Fundamentals of Differential Equations, Second Edition.*** R. Kent Nagle, Edward B. Saff. Benjamin/Cummings, 1989, xv + 711 pp, $43.95. [ISBN: 0-8053-0254-9] A text for the standard introductory ordinary differential equations course. Also includes one chapter on partial differential equations. Optional numerical techniques are introduced early, systems are treated briefly near the end. Each chapter ends with an interesting set of extended applications. (*First Edition*, TR, May 1987; Extended Review, December 1987.) SP

**Differential Equations, S(17-18), P.** ***Handbook of Differential Equations.*** Daniel Zwillinger. Academic Pr, 1989, xx + 673 pp, $49.95. [ISBN: 0-12-784390-6] A clearly-written, well-organised, comprehensive reference work collecting nearly 200 methods for solving ordinary differential equations and partial differential equations. Four chief parts: definitions and concepts; exact methods; approximate analytical methods; numerical methods. Each method is described in a short section, typically with pertinent references, brief examples, important theorems, computer techniques, etc. In summary, a remarkable guidebook for a vast territory. PZ

**Differential Equations, T(14-15: 1).** ***Introduction to Ordinary Differential Equations, Fourth Edition.*** Shepley L. Ross. Wiley, 1989, xi + 609 pp, $48.16. [ISBN: 0-471-09881-7] A new edition of a well-known textbook. Additions include sections on the use of matrices to solve linear systems with constant coefficients and expanded treatment of numerical methods. Leibnis notation for derivatives has been replaced with prime notation. 360 new exercises. (*First Edition*, TR, February 1967; *Second*

*Edition*, TR, May 1975; Extended Review, November 1976; *Third Edition*, TR, January 1981.) SM

**Differential Equations, S(14-16), C.** *IBM PC. Differential and Difference Equations through Computer Experiments, Second Edition.* Hüseyin Koçak. Springer-Verlag, 1989, xvii + 224 pp, $49.95 (P). [ISBN: 0-387-96918-7] Version 1.1 of the very popular "Phaser" simulation package for IBM personal computers. This version uses either the high resolution EGA or VGA graphics or the (original) lower resolution CGA graphics. The software plots trajectories of differential and difference equations. The manual provides a synopsis of the relevant mathematics, a handbook of Phaser commands, and a library of standard examples of one-, two-, and three-dimensional equations that in themselves constitute an outline of elementary differential equations. LAS

**Partial Differential Equations, P.** *Integral Manifolds and Inertial Manifolds for Dissipative Partial Differential Equations.* P. Constantin, *et al.* Appl. Math. Ser., V. 70. Springer-Verlag, 1989, x + 123 pp, $34. [ISBN: 0-387-96729-X] An inertial manifold is an invariant manifold that attracts exponentially all trajectories of a partial differential equation. This study focuses on some features of inertial manifolds that reveal similarities between the long term evolution of solutions of dissipative partial differential equations and solutions of finite dimensional dynamical systems. LAS

**Partial Differential Equations, P.** *Multigrid Methods: Theory, Applications, and Supercomputing.* Ed: S.F. McCormick. Lect. Notes in Pure & Appl. Math., V. 110. Marcel Dekker, 1988, xiv + 641 pp, (P). [ISBN: 0-8247-7979-7] Thirty-three research papers presented at the Third Copper Mountain Conference on Multigrid Methods held April 5-10, 1987 at Copper Mountain, Colorado. AO

**Partial Differential Equations, P.** *Recent Developments in Hyperbolic Equations.* Ed: L. Cattabriga, *et al.* Pitman Res. Notes in Math. Ser., V. 183. Longman Scientific & Technical (US Distr: Wiley), 1988, 411 pp, $64.95 (P). [ISBN: 0-582-03491-4] Papers from a semester conference at the University of Pisa, January-June, 1987. Topics include propagation of singularities, well-posed weakly hyperbolic Cauchy problems, solvability in Gevrey and analytic classes, global solutions for nonlinear equations with small initial data, nonstrictly hyperbolic systems, and ramification of solutions to equations of Kowalewsky type. MR

**Partial Differential Equations, P.** *Asymptotic Analysis of a Class of Perturbed Korteweg-de Vries Initial Value Problems.* F. de Kerf. CWI Tract, V. 50. Math Centrum, 1988, 180 pp, Dfl. 26.70 (P). [ISBN: 90-6196-351-6] Certain solutions of the Korteweg-de-Vries equation $u_t - 6uu_x + u_{xxx} = 0$ are solitons—solitary waves that exhibit particle-like behavior when they interact. This monograph examines the behavior of solutions under slight perturbations of the original equations as a means of studying the effect of models that neglect small terms. LAS

**Partial Differential Equations, P.** *Nonlinear Partial Differential Equations and their Applications, Collège de France Seminar, Volume IX.* Ed: H. Bresis, J.L. Lions. Pitman Res. Notes in Math. Ser., V. 181. Longman Scientific & Technical (US Distr: Wiley), 1988, 375 pp, $64.95 (P). [ISBN: 0-582-02181-2] Thirteen papers on nonlinear partial differential equations, with applications to plasticity, liquid crystals, fiber optics (a long chapter), quantum mechanics, and even economics. BC

**Partial Differential Equations, P.** *Mathematical Analysis of Nonlinear Dynamic Processes.* K-U Grusa. Pitman Res. Notes in Math. Ser., V. 176. Longman Scientific & Technical (US Distr: Wiley), 1988, 450 pp, $79.95 (P). [ISBN: 0-582-02880-9] A detailed introduction to the modelling of nonlinear phenomena by partial differential equations on multi-connected domains, and controlled pointwise or by game theoretic strategies. Models considered in depth include population migration (reaction-diffusion equations), new influence phenomena in economics (nonlinear wave equations), and cardiac arhythmias (vortical processes). CE

**Partial Differential Equations, P.** *Calculus on Heisenberg Manifolds.* Richard Beals, Peter Greiner. Annals of Math. Stud., No. 119. Princeton U Pr, 1988, ix + 194 pp, $15.95 (P); $40. [ISBN: 0-691-08501-3; 0-691-08500-5] Studies certain pseudodifferential operators on smooth manifolds. Topics include operators and their models, composition of operators, hypoellipticity and parametrics for second-order operators. GN

**Partial Differential Equations, T(15-16), L.** *Partial Differential Equations: Theory and Technique, Second Edition.* George F. Carrier, Carl E. Pearson. Academic Pr, 1988, xi + 340 pp, $39.95. [ISBN: 0-12-160451-9] Still an "attractive introduction to the subject" (*First Edition*, TR, June-July 1976). Changes from that edition include an added chapter on transform methods, a section on integral equations, and additional exercises. GN

**Numerical Analysis, T(15-16).** *Numerical Analysis, Fourth Edition.* Richard L. Burden, J. Douglas Faires. PWS-Kent, 1989, xv + 729 pp. [ISBN: 0-53491-585-X] Intended for a junior/senior year course (1 or 2 terms), this book aims at science and engineering students who know Basic, Fortran, or Pascal. Changes from *Third Edition* include stronger emphasis on factorisation of matrices, more attention to norms commonly applied to vectors and matrices, greater emphasis on the power method of approximating eigenvalues, and (of course) more exercises (some modeled after problems on recent exams of the Society of Actuaries). (*First Edition*, TR, November 1978; Extended Review, March 1980.) AWR

**Numerical Analysis, S(16-17).** *Schaum's Outline of Theory and Problems of Numerical Analysis, Second Edition.* Francis Scheid. McGraw-Hill, 1989, 471 pp, $11.95 (P). [ISBN: 0-07-055221-5] Pretty much what one would expect from a Schaum's Outline: an organised and very concise treatment of un-

dergraduate numerical analysis. Hundreds of solved and unsolved problems. The author writes in a lively and very enjoyable style. SM

**Numerical Analysis, P.** *Iterative Methods for the Solution of Linear Systems.* Ed: A. Hadjidimos. North-Holland (US Distr: Elsevier Science), 1988, 291 pp, $78. [ISBN: 0-444-87277-9] A collection of twenty invited papers intended to sample trends in research in this area, and to provide a broad overview of the field as it is developing. AWR

**Numerical Analysis, T(14-15: 1, 2), L.** *Applied Numerical Analysis, Fourth Edition.* Curtis F. Gerald, Patrick O. Wheatley. Addison-Wesley, 1989, xi + 722 pp, $40.75. [ISBN: 0-201-11583-2] Several chapters have been reorganised and rewritten for greater clarity and to bring the material more up-to-date. This edition features more emphasis on the use of matrix algebra, an elementary introduction to finite-element methods, and two-color printing. (*First Edition,* TR, August-September 1970; *Second Edition,* TR, November 1978; *Third Edition,* TR, January 1985.) AO

**Numerical Analysis, T(15-16: 1).** *Introduction to Applied Numerical Analysis.* Richard W. Hamming. Hemisphere Pub, 1989, x + 331 pp, $29.95. [ISBN: 0-89116-865-6] A reprint of the original edition published by McGraw-Hill in 1971 (Extended Review, April 1977). No significant changes have been made to the text. AO

**Numerical Analysis, C.** *Basic Numerical Modules in Fortran for the IBM PC and Compatibles.* Karol Bõning. Friedr. Vieweg & Sohn, 1988, iii + 320 pp, DM 298 (P). [ISBN: 3-528-02793-2] Software package for getting numerical results. 117 modules on three disks. For each, the accompanying documentation identifies the purpose and method; external subroutines, data and parameters used; results obtained; size. Not a text; few of the methods are derived or explained. Topics include elementary calculations, interpolation, approximation, solving linear equations, eigen problems, nonlinear equations, numerical integration, ordinary and partial differential equations. DFA

**Functional Analysis, T(17-18), P, L.** *Mathematical Analysis and Numerical Methods for Science and Technology, Volume 2: Functional and Variational Methods.* Robert Dautray, Jacques-Louis Lions. Transl: Ian N. Sneddon. Springer-Verlag, 1988, xv + 561 pp, $89.90. [ISBN: 0-387-19045-7] Translation of Chapters III to VII of *Analyse mathématique et calcul numérique pour les sciences et les techniques.* Each chapter is written by different author teams. Part of a 21-chapter, six-volume treatise sponsored by the French atomic energy commission. Topics: functional transformations, Sobolev spaces, linear differential operators, operators on Banach and Hilbert spaces, linear variational problems, distribution theory. LAS

**Functional Analysis, P.** *Schauder Bases: Behaviour and Stability.* P.K. Kamthan, M. Gupta. Pitman Mono. & Surv. in Pure & Appl. Math., V. 42. Longman Scientific & Technical (US Distr: Wiley), 1988, 514 pp, $79.75. [ISBN: 0-470-21029-X] This book is intended by the authors to serve as the completion of their work on Schauder bases and contains a good part of the recent results in this area. It assumes a comfortable familiarity with the theory of locally convex topological vector spaces. Part I is a summary of the results needed as background; later results are more technical and detailed. Extensive bibliography; index. Excessive use of abbreviations. JS

**Analysis, S*(16-18), P*, L.** *A Handbook of Fourier Theorems.* D.C. Champeney. Cambridge U Pr, 1989, xi + 185 pp, $17.95 (P). [ISBN: 0-521-36688-7] A brief, clear summary of Fourier analysis: basics of real analysis (including a brief introduction to Lebesgue integration and $L^p$ theory), convergence and inversion theorems, and elementary distribution theory. Pitched at late undergraduate to early graduate level, but should appeal to any working mathematician or scientist who wants an "aerial" view of the subject. Nothing is proved; much is explained. (1987 hardcover edition, TR, August-September 1988.) PZ

**Analysis, P.** *Lecture Notes in Mathematics-1336: Semi-Classical Analysis for the Schrödinger Operator and Applications.* Bernard Helffer. Springer-Verlag, 1988, v + 107 pp, $13.10 (P). [ISBN: 0-387-50076-6] "The purpose of the semi-classical analysis is to understand, from a mathematical point of view, the general correspondence principle of the Quantum mechanics saying that, when the Planck constant $h$ tends to zero, we must recover starting from the Quantum mechanics, the classical mechanics." Develops theory of tunneling effect and applies to Schrödinger operators with periodic electric potentials and Schrödinger operators with magnetic fields. BH

**Analysis, T*(17-18: 2), L.** *Principles of Applied Mathematics: Transformation and Approximation.* James P. Keener. Addison-Wesley, 1988, xv + 560 pp, $48.50. [ISBN: 0-201-15674-1] For a course in applied mathematical techniques. Engaging style. Develops and applies spectral theory of operators and asymptotic analysis to a wide range of problems. Chapter subjects: finite dimensional vector spaces, function spaces, integral equations, differential operators, calculus of variations, complex variable theory, transform and spectral theory, partial differential equations, inverse scattering transform, asymptotic expansions, regular and singular perturbation theory. Many real examples, exercises, references; some discussion of numerical packages. DFA

**Analysis, S(17-18), P.** *Harmonic Analysis in Phase Space.* Gerald B. Folland. Annals of Math. Stud., No. 122. Princeton U Pr, 1989, ix + 277 pp, $17.50 (P); $55. [ISBN: 0-691-08528-5; 0-691-08527-7] A clearly-written treatment of the Heisenberg group, quantisation, the Weyl calculus of pseudodifferential operators, the metaplectic representation, wave packets, and other stuff that shows up in

Fourier analysis, partial differential equations, mathematical physics, representation theory, and number theory. BC

**Analysis, T(17: 2).** *Real Analysis and Probability.* Richard M. Dudley. Math. Ser. Wadsworth, 1989, xi + 436 pp, $52.95. [ISBN: 0-534-10050-3] Assumes analysis at the level of Rudin's classic. First five chapters form a concise yet thorough treatment of measure and integration, followed by a discussion of convexity and generalised differentiation. The probability in the title is a short, intense treatment of limit laws, distributions, and continuous processes. Ample exercises and useful bibliographies follow each chapter. TAV

**Analysis, P.** *Treatise on Analysis, Volume VII.* J. Dieudonné. Transl: Laura Fainsilber. Pure & Appl. Math., V. 10. Academic Pr, 1988, xiv + 366 pp, $69.95. [ISBN: 0-12-215507-6] Translation of Part One (on Pseudodifferential Operators) of Chapter XXIII: Linear Functional Equations of Dieudonné's monumental *Éléments d'analyse.* The French original was published in 1978, ten years after the first chapter in the treatise. Includes references for all seven volumes, none more recent than 1976. LAS

**Analysis, T(16-17: 1), S, P, L.** *Classical Fourier Transforms.* Komaravolu Chandrasekharan. Universitext. Springer-Verlag, 1989, 172 pp, $29.50 (P). [ISBN: 0-387-50248-3] Nuts and bolts of Fourier transforms on $L_i(-\infty,\infty), i=1,2$, including Tauberian theorems, Heisenberg's inequality, and Hardy's interpolation formula. Ends with a chapter on Fourier-Stieltjes transforms, and a uniqueness theorem of Offord on $L_1(-\omega,\omega)$. BC

**Analysis, T(15-16: 1), S, L*.** *Theory of Discrete and Continuous Fourier Analysis.* H. Joseph Weaver. Wiley, 1989, xii + 307 pp, $39.95. [ISBN: 0-471-62872-7] A unified introduction to Fourier series, Fourier transforms, discrete Fourier transforms, and the fast Fourier transform, written so as to make clear appropriate parallels. Begins with review of necessary analysis (e.g., integration and distribution theory), and concludes with a treatment of discrete sampling of a continuous function (via a "comb" function) that links the discrete and continuous perspectives in a common, important application. LAS

**Algebraic Geometry, S(18), P.** *Conformal Geometry.* Ed: Ravi S. Kulkarni, Ulrich Pinkall. Aspects of Math., V. E12. Friedr. Vieweg & Sohn, 1988, vii + 236 pp, DM 48 (P). [ISBN: 3-528-08982-2] Eight long papers, in advanced expository style, on conformal geometry, from a 1985-86 seminar at the Max-Planck-Institut. Several chapters, especially the first, are introductory. Accessible to advanced graduate students as an introduction to research topics. PZ

**Algebraic Geometry, T(16-17), L.** *Undergraduate Algebraic Geometry.* Miles Reid. Math. Soc. Stud. Texts, V. 12. Cambridge U Pr, 1988, viii + 129 pp, $34.50; $12.95 (P). [ISBN: 0-521-35559-1; 0-521-35662-8] Let's say advanced undergraduates: Those familiar with commutative rings, principal ideal domain (PID's), Galois theory, some topology, and possessing mathematical sophistication. Given this, the author leads the student on a lively, interesting, down-to-earth tour of the fundamental algebraic geometry. The author covers conics, cubics, affine and projective varieties, the 27 lines on a cubic surface, and closes with some welcome, provocative comments on "modern" algebraic geometry. SG

**Differential Geometry, S(16-17).** *Schaum's Outline of Theory and Problems of Tensor Calculus.* David C. Kay. McGraw-Hill, 1988, 228 pp, $10.95. [ISBN: 0-07-033484-6] For both undergraduates and graduate students in need of mastering the basic concepts and methods of tensors. The traditional component approach (replete with subscripts and superscripts). Chapters 1 and 2 contain the requisite background material on summation convention and basic linear algebra for tensors. Later chapters on tensors in classical mechanics, in special relativity, and in Euclidean geometry. Final chapter features the modern noncomponent approach. Explanatory sections are clear and helpful and reflect the author's teaching experience with the material. Much elementary classical differential geometry is met along the way. JK

**Differential Geometry, T(18: 4), S, P.** *Lecture Notes in Mathematics-1353: Critical Point Theory and Submanifold Geometry.* Richard S. Palais, Chuu-lian Terng. Springer-Verlag, 1988, x + 271 pp, $24.30 (P). [ISBN: 0-387-50399-4] Part I is a modern introduction, suitable for graduate students, to submanifold geometry including submanifolds of Hilbert spaces. The goals are two-fold: to classify all isoparametric submanifolds, and to develop relationship between the topology and geometry of these manifolds. Part II is devoted to critical point theory of Hilbert manifolds. The two parts are connected using the Morse Index Theorem. Overall, this seems better written than the average entry in the Lecture Notes series. Includes exercises. MR

**Differential Geometry, P.** *Differential Geometry of Frame Bundles.* Luis A. Cordero, C.T.J. Dodson, Manuel de León. Math. & Its Applic. Kluwer Academic, 1989, x + 234 pp, $74. [ISBN: 0-7923-0012-2] Bundles are increasingly important both as a setting for global analysis on manifolds, and by virtue of their applications to gauge field theories and general relativity. This work details the differential geometry of frame bundles. The approach is coordinate free and via jets, and assumes some graduate level knowledge of differential geometry. CE

**Geometry, S.** *Bendings of Surfaces and Stability of Shells.* A.V. Pogorelov. Transl. of Math. Mono., V. 72. AMS, 1988, viii + 77 pp, $41. [ISBN: 0-8218-4525-X] This is a beautifully focused little book on the geometric theory of stability of elastic shells. It includes the complete solution of the problem of stability of shells under an extremal pressure with no assumptions regarding the character of buckling. AWR

**Algebraic Topology, P.** *An Introduction to Intersection Homology Theory.* Frances Kirwan. Pitman

Res. Notes in Math. Ser., V. 187. Longman Scientific & Technical (US Distr: Wiley), 1988, 169 pp, $47.95 (P). [ISBN: 0-470-21198-9] Based on lectures given in Oxford in 1987, these notes comprise a "beginner's guide" to intersection homology theory, a recent extension of the ordinary homology theory of manifolds, that gives better results for singular spaces such as complex projective varieties and topological pseudomanifolds. Not intended to be comprehensive, the notes emphasise motivating and explaining the main ideas and definitions. CE

**Differential Topology, P.** *Lecture Notes in Mathematics-1350: Differential Topology.* Ed: U. Koschorke. Springer-Verlag, 1988, vi + 269 pp, $24.30 (P). [ISBN: 0-387-50369-2] Proceedings of the Second Siegen Topology Symposium held July 27-August 1, 1987 at the University of Siegen, Federal Republic of Germany. Sixteen research articles grouped into three categories: linking phenomena and three-dimensional topology, immersions and vector bundle morphisms, and manifolds and algebraic topology. GG

**Topology, P*, L**.** *Elements of Mathematics: General Topology.* Nicolas Bourbaki. Springer-Verlag, 1989. *Chapters 1-4,* vii + 437 pp, $69 [ISBN: 0-387-19374-X]; *Chapters 5-10,* iv + 363 pp, $59. [ISBN: 0-387-19372-3] Reprint of Addison-Wesley's 1966 English translation of the famous French Bourbaki series which shaped so profoundly the nature of contemporary mathematics. Each chapter features, in addition to archetypal axiomatic development of theory, many very challenging exercises and superb historical notes (TR of 1966 edition, June 1967; Extended Review, December 1972). LAS

**Topology, P, L**.** *The Collected Papers of R.H. Bing.* Ed: Sukhjit Singh, Steve Armentrout, Robert J. Daverman. AMS, 1988, $155 set [ISBN: 0-8218-0117-1]. *Volume 1,* xix + 886 pp; *Volume 2,* xvii + 759 pp. A splendid celebration of Bing's mathematical life including all his mathematical papers and abstracts organised in part with Bing's help. Highlight is an autobiographical essay prepared by Singh from tapes made by Bing. Includes lists of Bing's Ph.D. students, family photographs, and a biography by R.D. Anderson and C.E. Burgess. LAS

**Optimisation, T(16-18: 1, 2), S, L.** *Simulated Annealing and Boltzmann Machines.* Emile Aarts, Jan Korst. Disc. Math. & Optimis. Wiley, 1989, xii + 272 pp, $49.75. [ISBN: 0-471-92146-7] Simulated annealing is an approach to large problems of combinatorial optimisation by direct analogy with the cooling process of annealing which minimises free energy—or cost—without becoming trapped in locally minimal configurations. A Boltsmann machine is a massively parallel computer—a neural network model—with two-state components (on or off) governed by a consensus function that measures the desirability of a global configuration. This monograph discusses implementation of simulated annealing algorithms on Boltsmann machines, where cost, free energy, and consensus play analogous role. LAS

**Optimisation, P.** *Lecture Notes in Mathematics-1354: Approximation and Optimization.* Ed: A. Gómes, *et al.* Springer-Verlag, 1988, vi + 280 pp, $25.70 (P). [ISBN: 0-387-50443-5] Contains the Proceedings of the Seminar on Approximation and Optimisation held in January 1987 at the University of Havana in Havana, Cuba. SM

**Optimisation, T(16-17: 1, 2), L.** *Practical Methods of Optimization, Second Edition.* R. Fletcher. Wiley, 1987, xiv + 436 pp, $53.95. [ISBN: 0-471-91547-5] Both volumes of the previous edition have been combined into a single book and new material added. Part 1 covers unconstrained optimisation, and Part 2 covers constrained optimisation. Emphasises methods known to be reliable and efficient. (TR, *First Edition, V. 1,* March, 1981; TR, *First Edition, V. 2,* November 1982.) AO

**Dynamical Systems, P.** *Chaos and Integrability in Nonlinear Dynamics, An Introduction.* Michael Tabor. Wiley, 1989, xiii + 364 pp, $55. [ISBN: 0-471-82728-2] An introductory account of nonlinear dynamics and chaos with an emphasis on Hamiltonian (non-dissipative) systems. Care is taken to place the new results in the context of traditional treatments of ordinary differential equations and classical mechanics. Topics include perturbation and KAM theory, area preserving mappings, chaos and integrability in the semi-classical limit, solitons and the KdV equations. A one-chapter introduction to chaos in dissipative systems is also given. CE

**Dynamical Systems, P.** *Charles Conley Memorial Volume: Special Issue of Ergodic Theory and Dynamical Systems, Volume 8.* Ed: M.R. Herman, *et al.* Cambridge U Pr, 1988, 409 pp, $69.50. [ISBN: 0-521-369290] A special issue of the journal *Ergodic Theory and Dynamical Systems* dedicated to the memory of Charles Conley. It includes a brief biographical sketch and a previously unpublished yet seminal work of Conley's on chain recurrent sets and Morse decomposition. The remaining twenty-one research articles range widely, and include contributions by many luminaries of dynamical systems theory. CE

**Modelling, T**(15).** *Mathematical Models and Their Analysis.* Frederic Y.M. Wan. Harper & Row, 1989, xvi + 394 pp. [ISBN: 0-06-046902-1] Outstanding text on not only mathematical models, but also on the process of mathematical modelling. The models studied are grouped according to the principle concern of the model. For example, under the heading of stability of equilbrium, the text looks at periodic orbits, hair (Euler buckling and elastic stability), and car following (Lagrangian and Eulerian formulation of traffic flow). Other models covered include fishing rates, population distributions, economic growth, precession of Murcury, and timber harvesting. Necessary background is a familiarity with differential equations. Many exercises. Highly recommended. MR

**Modelling, P*, L*.** *Alternate Realities: Mathematical Models of Nature and Man.* John L. Casti.

Wiley, 1989, xvii + 493 pp, $34.95. [ISBN: 0-471-61842-X] A bold, wide-ranging introduction to the use of modern mathematical tools (discrete dynamics, catastrophe theory, dynamical systems, game theory, optimal control theory) in the process of modelling natural and human systems. Uses a full array of graduate-level mathematical notation, often in innovative contexts. Avoids typical topics (queueing, regression, LP); applications include DNA sequences, cognitive states, power generation, metabolism, and sociobiology. Unorthodox, philosophical, and purposeful in offering a highly innovative "theory of models." LAS

**Modelling, T(15-17), S, L*.** *Empirical Model Building.* James R. Thompson. Appl. Prob. & Stat. Wiley, 1989, xiv + 242 pp, $39.95. [ISBN: 0-471-60105-5] A wide-ranging collection of problem situations suitable for use in undergraduate modelling courses (annuities, demographics of ancient Israel, models for AIDS) together with a miscellany of tools (e.g., exploratory data analysis, Arrow's and Stein's paradoxes, quality control) useful to the practicing modeller. A useful core text, especially if supplemented by real problems reflecting local context. LAS

**Control Theory, T(17: 2), P.** *Control Systems Synthesis: A Factorization Approach.* M. Vidyasagar. Ser. in Signal Process., Optimis., and Control, V. 7. MIT Pr, 1987, 436 pp, $15 (P); $40. [ISBN: 0-262-22027-X] This is an account of work done largely over the last five years that uses algebraic structures as a framework in which to study multivariable feedback control systems. Pioneered by the author, the method relies on "factoring" the transfer matrix of a not necessarily stable system as the "ratio" of two stable rational matrices. Three appendices survey the mathematical prerequisites from abstract algebra, matrices, and topological rings. Chapter 3 outlines the method in the context of single-input-single-output systems. Attractively written for researchers or for a graduate text. AWR

**Control Theory, P.** *Optimization and Identification of Systems Governed by Evolution Equations on Banach Space.* N.U. Ahmed. Pitman Res. Notes in Math. Ser., V. 184. Longman Scientific & Technical (US Distr: Wiley), 1988, 187 pp, $49.95 (P). [ISBN: 0-470-21171-7] Topics covered include existence and uniqueness conditions for parabolic and hyperbolic systems, linear and nonlinear semigroups, boundary value problems, and filtering of stochastic evolution equations. The control parameters in question are operators which may be bounded, relatively bounded, or unbounded. Includes a chapter on numerical results. MR

**Systems Theory, P, L.** *Quality Control, Robust Design, and the Taguchi Method.* Ed: Khosrow Dehnad. Stat./Prob. Ser. Wadsworth, 1989, xxiv + 309 pp, $42.95. [ISBN: 0-534-09048-6] A collection of the most highly sought papers produced by members of AT&T Bell Laboratories Quality Assurance Center who have worked with Genichi Taguchi. The papers are organised into three parts: an overview, case studies, and methodology. SM

**Probability, S(15-17).** *Exercises in Probability.* T. Cacoullos. Springer-Verlag, 1989, ix + 248 pp, $49.80. [ISBN: 0-387-96735-4] Primarily a collection of 329 problems, with complete solutions, ranging from straightforward to challenging. Also includes brief summaries of the basic theory and formulas, and supplements containing additional problems with only answers or hints. Does not include stochastic processes. RSK

**Probability, P.** *Estimating and Choosing: An Essay on Probability in Practice.* Georges Matheron. Transl: A.M. Hasofer. Springer-Verlag, 1989, ix + 141 pp, $39.80 (P). [ISBN: 0-387-50087-1] A translation of Matheron's work *Estimer et choisir,* written in 1978. Presents a method for the use of probabilistic models to describe unique phenomena in a purely objective way. It is written with geostatistical applications in mind, but the method is applicable to any probabilistic modelling. RH

**Probability, P.** *Algebraic Probability Theory.* Imre Z. Russa, Gábor J. Székely. Prob. & Math. Stat. Wiley, 1988, xii + 251 pp, $69.95. [ISBN: 0-471-91803-2] Random variables imply distributions, sums of random variables imply convolutions of the distributions, and this operation on the set of distributions yields a semigroup. Many results in probability follow from the properties of topological semigroups. The approach, while technical, yields some new insights. Contains an extensive annotated bibliography. TAV

**Stochastic Processes, T(17: 1), P, L.** *Stopped Random Walks: Limit Theorems and Applications.* Allan Gut. Appl. Prob., V. 5. Springer-Verlag, 1988, ix + 199 pp, $44. [ISBN: 0-387-96590-4] Most treatments of random walks look at the state of the walk after a fixed number of steps. More natural is to consider stopping the walk after a random number of steps—the theme of this text. A very readable treatment of a fascinating subject. Substantial bibliography; no exercises. TAV

**Stochastic Processes, P*.** *Introduction to the Theory of Coverage Processes.* Peter Hall. Prob. & Math. Stat. Wiley, 1988, xviii + 408 pp, $39.95. [ISBN: 0-471-85702-5] Roughly, a coverage process is a countable collection of sets in $R^n$ whose centers form a (spatial) point process. Theory and techniques are presented in detail. The writing is clear and carries the reader through significant mathematical sophistication. TAV

**Stochastic Processes, T*(14: 1), P, L.** *Random Processes, A First Look, Second Edition Revised and Expanded.* R. Syski. Stat: Textbooks & Mono., V. 98. Marcel Dekker, 1989, xxii + 413 pp, $49.75. [ISBN: 0-8247-8028-0] Revised version; new material includes chapter on statistical estimation and sections on birth-death and branching processes (*First Edition,* TR, March 1980). Intuitive presentation of stochastic processes and some probability. Calculus prerequisite. TH

**Stochastic Processes, T*(15: 2), L.** *Quantitative Forecasting Methods.* Nicholas R. Farnum, LaVerne W. Stanton. PWS-Kent, 1989, xvi + 573 pp. [ISBN: 0-534-91686-4] Readable introduction at intermediate level; emphasises applications. Models for series with and without trend, seasonal, cyclical. Separate chapters for regression and Box Jenkins (no transfer function) models. Model building process is clear. Perhaps over-reliance on statistical tests rather than on judgment in choosing models and terms. TH

**Stochastic Processes, T(17: 1).** *Point Process Models with Applications to Safety and Reliability.* W.A. Thompson, Jr. Chapman & Hall, 1988, xi + 146 pp, $39.95. [ISBN: 0-412-01481-5] Monograph dealing with models for the placement of points on the positive time axis according to some chance mechanism. Presumes a background of a rigorous course in probability theory. Main application to safety is the safety of nuclear power plants. RSK

**Elementary Statistics, T*(13-17: 1).** *Introduction to the Practice of Statistics.* David S. Moore, George P. McCabe. WH Freeman, 1989, xix + 839 pp, $38.95. [ISBN: 0-7167-1989-4] An elementary "but serious" text designed to help mathematically weak students think about data and understand statistical reasoning. Real data and real problems dominate this attractive text. Four chapters on data lead via simple probability to four chapters on inference. Written to accompany a new 26-program PBS Telecourse *Against All Odds* developed by David Moore and produced by COMAP, but can easily be used without the video support. A MINITAB computer supplement is available. LAS

**Elementary Statistics, T(15-17), S.** *Statistical Reasoning in Law and Public Policy, Volume 2: Tort Law, Evidence, and Health.* Joseph L. Gastwirth. Stat. Modeling & Decision Sci. Academic Pr, 1988, viii + 487 pp, $84.50. [ISBN: 0-12-277161-3] Second volume (TR, *Volume 1*, March 1989) of an innovative monograph that teaches basic statistics in the context of real case law. This volume concentrates on survey evidence, small probabilities (in disease rates), Bayesian inference, medical studies, and epidemiologic evidence. LAS

**Statistics, P.** *Computational Aspects of Survey Data Processing.* L.C.R.J. Willenborg. CWI Tract, V. 54. Math Centrum, 1988, 154 pp, Dfl. 24.20 (P). [ISBN: 90-6196-356-7] From Ph.D. thesis describing research project with Netherlands Central Bureau of Statistics on computerising collection and analysis of survey data. Describing and testing logical structure of questionnaires, data editing, imputation of missing values. TH

**Statistics, P, L.** *Analog Estimation Methods in Econometrics.* Charles F. Manski. Mono. on Stat. & Appl. Prob. Chapman & Hall, 1988, xv + 158 pp, $45. [ISBN: 0-412-01141-7] The sample average is an analog estimate for the population mean. Develops the analogy principle for use in estimation, regression, moments, conditional likelihood, and asymptotic theory for moment estimation. Econometrics is not a central theme. TH

**Statistics, P.** *Lecture Notes in Statistics-49: Extremal Families and Systems of Sufficient Statistics.* Steffen L. Lauritsen. Springer-Verlag, 1988, xv + 268 pp, $28 (P). [ISBN: 0-387-96872-5] Revised version of the author's 1982 *Statistical Models as Extremal Families* (errors corrected, minor additions). Given that one wants to perform a certain statistical analysis, how can this be expressed in terms of a statistical model? Specified repetitive structure of an experiment and sufficient statistics with known conditional distributions lead to extremal families; this is an alternate way of deriving well-known and new statistical models. TH

**Statistics, P, L.** *The Evolving Role of Statistical Assessments as Evidence in the Courts.* Ed: Stephen E. Fienberg. Springer-Verlag, 1989, xvii + 357 pp, $34. [ISBN: 0-387-96914-4] An appraisal of forms in which statistical assessments are presented and admitted into evidence in federal and state courts, including recommendations to improve comprehension of statistical evidence by judges and juries. Includes case studies,. expert testimony, and reviews of discrimination, antitrust, and environmental litigation; contains extensive, useful appendices. Report of a Panel on Statistical Assessments as Evidence in the Courts of the Commission on Behavioral and Social Sciences of the National Research Council. LAS

**Statistics, P*.** *Statistical Inference for Spatial Processes.* B.D. Ripley. Cambridge U Pr, 1988, viii + 148 pp, $34.50. [ISBN: 0-521-35234-7] Prise winning essay dealing with some of the distinctive features and special problems of statistical inference on spatial stochastic processes, such as the importance of edge effects and the lack of a unique asymptotic setting. Final half is devoted to concerns of digital image processing. RSK

**Statistics, S.** *Statistical Tables and Formulae.* Stephen Kokoska, Christopher Nevison. Texts in Stat. Springer-Verlag, 1989, 88 pp, $9.95 (P). [ISBN: 0-387-96873-3] Concise collection of probability and statistics formulas, comprising more than half the book, together with a collection of standard tables, including tables for nonparametric tests. RSK

**Statistics, S(16-17).** *Problem Solving: A Statistician's Guide.* Christopher Chatfield. Chapman & Hall, 1988, xvi + 261 pp, $35 (P); $79.95. [ISBN: 0-412-28680-7; 0-412-28670-X] Practical guide on how to handle statistical problems. First part provides general principles, reflecting the author's experiences. Second part presents a series of problem-oriented exercises, with 'solutions,' illustrating the practical problems of real data analysis. Final part includes a summary of techniques assumed to have been previously studied. Note price. RSK

**Statistics, T(16-17: 1), P.** *The Method of Paired Comparisons, Second Edition, Revised.* H.A. David. Griffin's Stat. Mono. & Courses, V. 41. Oxford U Pr, 1988, 188 pp, $39.95. [ISBN: 0-85264-290-3] Revision of the author's 1963 *First Edition.* Treats the methodology of handling experiments in which ob-

jects to be compared are presented in pairs to be judged. Emphasis is on nonparametric procedures. Good set of references. RSK

**Statistics, P*.** *Lecture Notes in Statistics-47: {2}-Inverses and Their Statistical Application.* Albert J. Getson, Francis C. Hsuan. Springer-Verlag, 1988, viii + 110 pp, $16.30 (P). [ISBN: 0-387-96849-0] A {2}-inverse for a given matrix $A$ is any matrix $G$ satisfying $GAG = G$. (The more familiar {1}-inverse or $g$-inverse satisfies $AGA = A$.) This monograph provides a comprehensive study of the geometric characterisation, algebraic properties, and uses in statistics of the {2}-inverse. One result is a simplification of matrix expressions used in linear models. RSK

**Statistics, S**(13-16), L**.** *Statistics: A Guide to the Unknown, Third Edition.* Ed: Judith M. Tanur, *et al.* Wadsworth, 1989, xxv + 284 pp, $16.95 (P). [ISBN: 0-534-09492-9] An enduring classic intended to make clear to the public the types of contributions that mathematicians make to society, explained without the use of mathematics. Twenty-nine short essays illustrate the role of statistics in the biological, political, social, and physical worlds. Twelve new essays have been added in this edition, and the original essays have been updated by their authors. First published in 1972 (*First Edition,* TR, January 1973; Extended Review, April 1974; *Second Edition,* TR, March 1979) by the ASA-NCTM Joint Committee on Statistics and Probability. LAS

**Statistics, P, L*.** *Symposium on Statistics in Science, Industry, and Public Policy.* Board on Mathematical Sciences, 1989, vii + 53 pp, (P). Typescript record of an April 9, 1987 symposium at the National Academy of Sciences featuring four presentations and subsequent discussion on topics related to industry, to computing, to public policy, and to AIDS. Addressed to Washington policy makers; very readable. LAS

**Computer Literacy, T(13: 1), S.** *Computers Today, Third Edition.* Donald H. Sanders. McGraw-Hill, 1988, xxix + 640 pp, $33.95 (P). [ISBN: 0-07-054854-4] Workbook with words, words, words: about computers (their capabilities, their limitations, their impact), about terminology and computer lingo (what is a floppy disk? what is a smart terminal?), and about using computers and computer software. All modules from previous edition (*Second Edition,* TR, February 1986) have been dismantled, reworked, reassembled, with fresh material. Also available: study guides and instructor resource kit. Mathematical content very limited. LCL

**Elementary Computer Science, S*, C, L**.** *The Armchair Universe: An Exploration of Computer Worlds.* A.K. Dewdney. WH Freeman, 1988, xiii + 330 pp, $15.99. [ISBN: 0-716-71938-X] Columns of "Computer Recreations" reprinted, with addenda, from the *Scientific American* column from 1984 through 1987. Columns grouped into seven chapters: fractals, gadgets (e.g., hypercubes), artificial intelligence, cellular automata, word puzzles, simulation, and "core wars." Each chapter contains sufficient detail on algorithms to allow most amateur computer buffs to implement them in any language on any machine. Appendices give bibliographies for further reading and sources of software for those who would rather purchase than program. Altogether, a pleasant and very informal introduction to a significant subset of computer science. LAS

**Elementary Computer Science, T(13: 1).** *Computer Science 2: Principles of Software Engineering, Data Types, and Algorithms.* Henry M. Walker. Scott Foresman, 1989, xvii + 637 pp, $25 (P). [ISBN: 0-673-39829-3] A textbook for an updated CS2 course that emphasises software engineering principles and techniques, and exposes students to the breadth of the discipline of computer science. Pascal is used for the examples and in some of the exercises. AO

**Elementary Computer Science, T*(13: 1, 2).** *TURBO Pascal: An Introduction to the Art and Science of Programming, Second Edition.* Walter J. Savitch. Ser. in Struct. Program. Benjamin Cummings, 1988, xxvi + 803 pp, $28.95. [ISBN: 0-8053-8396-4] For the CS1 course. No prerequisites. TURBO Version 3.0 form of the second edition of the author's standard Pascal text; comes with *Version 4.0 Supplement.* Optional sections identify differences in standard Pascal. Some attention to CS2 topics: procedural and data abstraction, sorting and searching, hashing, numeric programming, loop invariants, dynamic data structures. Many examples, case studies, self-test, and programming exercises. DFA

**Programming, T, S, P*, L.** *The AWK Programming Language.* Alfred V. Aho, Brian W. Kernighan, Peter J. Weinberger. Addison-Wesley, 1988, x + 210 pp, (P). [ISBN: 0-201-07981-X] The first handbook on "awk," the immensely useful high-level UNIX text processing language. Each line in an awk program specifies a pattern with associated action to be taken if a line of text matches the pattern; input, fields, and common data types are handled automatically. Begins with a tutorial; includes a complete language definition, and many extended examples. LAS

**Programming, P, L.** *Object-Oriented Programming in Common Lisp: A Programmer's Guide to CLOS.* Sonya E. Keene. Addison-Wesley, 1989, xix + 266 pp, $27.95 (P). [ISBN: 0-201-17589-4] The language CLOS was developed by the X3J13 committee in its endeavor to develop a LISP standard. This book is both a tutorial on object-oriented programming in general, and a user's guided for CLOS. It is not a language specification for CLOS. SM

**Programming, T(15: 1), P, L.** *C as a Second Language: For Native Speakers of Pascal.* Tomasz Müldner, Peter W. Steele. Addison-Wesley, 1988, xiii + 575 pp, $29.25 (P). [ISBN: 0-201-19210-1] The authors provide an easy way for accomplished Pascal programmers to learn C via self study. The Kernighan and Ritchie standard is used throughout, with some discussions of extensions likely to be part of the ANSI C standard. The final chapter concerns

common C compilers available on Unix, MS-DOS, and Macintosh systems. SM

**Programming, T(13), S.** *Fortran 77 PDQ, Second Edition.* Thomas A. Boyle. Brooks/Cole, 1989, ix + 140 pp, $13 (P). [ISBN: 0-534-09936-X] No prerequisites. Informal and readable. For a short course or self-study. Covers the language quickly, from a problem-solving perspective. Examples and exercises. Handy reference guide to Fortran commands in an appendix. DFA

**Programming, T(13: 1, 2).** *LISP, Third Edition.* Patrick Henry Winston, Berthold Klaus Paul Horn. Addison-Wesley, 1989, xxi + 611 pp, $34.50 (P). [ISBN: 0-201-08319-1] Complete introduction to Common Lisp and some of its applications. Along with other new topics, this edition adds discussion of the Common Lisp Object System and techniques enabled by object-oriented programming, as well as of constraint propagation, backward chaining, key ideas in PROLOG. Has improved application examples and new ones on probability bounds, project simulation, visual object recognition. More emphasis on procedure and data abstraction; techniques for better procedure definition. (*First Edition,* TR, August-September 1981.) DFA

**Programming, T(13: 1), L.** *LISP: From Foundations to Applications.* G.I. Doukidis, V.P. Shah, M.C. Angelides. Chartwell-Bratt, 1988, 226 pp, (P). [ISBN: 0-86238-191-6] An introduction to muLISP which assumes no programming experience (although some, in whatever language, is helpful). Takes the reader on a rigorous path from elementary concepts to advanced techniques. Concludes with chapters on searching and knowledge representation (using frames). Many examples, exercises, and detailed solutions throughout. DFA

**Languages, T(15: 1).** *Comparative Programming Languages.* Leslie B. Wilson, Robert G. Clark. Intern. Comput. Sci. Ser. Addison-Wesley, 1988, xiii + 379 pp, $35.50. [ISBN: 0-201-18483-4] Designed to be used as the textbook for a modern "Organisation of Programming Languages" course. Rather than treating a series of languages, each chapter treats a single concept and examines the treatment of that concept in several languages—traditional imperative languages as well as functional, logic, and object-oriented languages. AO

**Algorithms, S(15-17), L.** *A Method of Programming.* Edsger W. Dijkstra, W.H.J. Feijen. Transl: Joke Sterringa. Addison-Wesley, 1988, viii + 188 pp, $29.25 (P). [ISBN: 0-201-17536-3] Presents a formal approach to program design in which a proof of a program's correctness is developed concurrently with the program. Several examples are given that illustrate the method. AO

**Algorithms, T(16-17: 1).** *Parallel Program Design: A Foundation.* K. Mani Chandy, Jayadev Misra. Addison-Wesley, 1988, xxviii + 516 pp, $39.75. [ISBN: 0-201-05866-9] Presents a theory of programming (a model of computation and an associated proof system) that allows programs to be systematically developed for a variety of architectures. Particular attention is given to dealing with problems of parallelism. AO

**Algorithms, S(17-18), P, L.** *Theoretical and Computational Aspects of Simulated Annealing.* P.J.M. van Laarhoven. CWI Tract, V. 51. Math Centrum, 1988, viii + 168 pp, Dfl. 25,30 (P). [ISBN: 90-6196-352-4] A doctoral dissertation giving details of implementation together with supporting computational evidence for approximating minimal configurations in hard combinatorial problems such as travelling salesman and job scheduling. Simulated annealing methods employ the Metropolis algorithm in analogy with the process of reaching thermodynamic equilibrium in annealing. In contrast to deterministic algorithms that find true minima, algorithms of this type converge non-monotonically to approximate solutions, accepting with decreasing probability transitions that increase costs. LAS

**Computer Systems, L.** *UNIX System V/386, Release 3.2.* AT&T. Prentice-Hall, 1989, (P). *User's Guide,* xxi + 663 pp [ISBN: 0-13-944869-1]; *System Administrators Reference Manual,* (P). [ISBN: 0-13-944950-7] The *User's Guide* provides a clear introduction to UNIX V/386 for novices, including tutorials on editors, shells, and communication programs. Appendices contain useful summaries. The *Reference Manual* contains Section 1 and 7 of the UNIX manual pages, covering all regular commands and I/O files. A permuted word index helps the reader locate appropriate pages. LAS

**Computer Systems, P.** *Parallel Processing for Scientific Computing.* Ed: Garry Rodrigue. SIAM, 1989, 428 pp, (P). [ISBN: 0-89871-228-9] Proceedings of the Third SIAM Conference on the title subject held in Los Angeles, December 1-4, 1987. Emphasis on numerical algorithms for parallel computing, but some of the 90 papers and abstracts concern scientific programming languages and environments, new computing architectures, parallel computer performance evaluation, applications to various scientific fields. DFA

**Computer Systems, P.** *Algorithm Animation.* Marc H. Brown. ACM Dist. Dissertat. MIT Pr, 1988, 186 pp, $30. [ISBN: 0-262-02278-8] Describes the development of BALSA-I and BALSA-II, environments for creating and interacting with animations of algorithms. The work described was done as part of the Electronic Classroom project at Brown University. AO

**Computer Graphics, T(16-17: 1), P, L.** *Curves and Surfaces for Computer Aided Geometric Design: A Practical Guide.* Gerald Farin. Comp. Sci. & Scient. Comput. Academic Pr, 1988, xv + 334 pp, $39.95. [ISBN: 0-12-249050-9] A unified treatment of the main ideas used in computer-aided geometric design. The main topics covered are Bésier and *B*-spline curves, rational Bésier and *B*-spline curves, geometric continuity, spline interpolation, and Coons patches. Presumes only a background in calculus and basic linear algebra. Two chapters are devoted

to the concepts from differential geometry that are needed. AO

**Theory of Computation, T(16-18: 1), P.** *Logic and Computation.* Lawrence C. Paulson. Tracts in Theoret. Comput. Sci., V. 2. Cambridge U Pr, 1987, xiii + 302 pp, $34.50. [ISBN: 0-521-34632-0] Introduction to Cambridge LCF, Logic for Computable Functions, an interactive theorem prover for reasoning about computable functions. First part outlines elementary logic and domain theory, the mathematics underlying LCF. Second part provides reference manual for Cambridge LCF. KS

**Artificial Intelligence, P, L.** *Machine Intelligence 11: Logic and the Acquisition of Knowledge.* Ed: J.E. Hayes, D. Michie, J. Richards. Clarendon Pr, 1988, viii + 460 pp, $85. [ISBN: 0-19-853718-2] Twenty papers from the Eleventh International Machine Intelligence Workshop held at Ross Priory, Scotland in 1985. Both theoretical papers and practical applications are contained in this work. SM

**Artificial Intelligence, S, P, L*.** *The Artificial Intelligence Debate: False Starts, Real Foundations.* Ed: Stephen R. Graubard. MIT Pr, 1988, 311 pp, $9.95 (P). [ISBN: 0-262-57074-2] Reprint of the winter 1988 issue of *Daedalus* featuring fourteen papers by authors such as Seymour Papert, Jack Cowan, Hilary Putnam, Jacob Schwartz, and Pamela McCorduck defining and debating the prospects of artificial intelligence, reflecting on unfulfilled promises, and speculating on promising directions for future research. Special focus is on "connectionism," the new hope of artificial intelligence based on models of the brain as a massively parallel computer in which memory is distributed as patterns among the whole rather than concentrated in individual neurons. LAS

**Artificial Intelligence, T(17), S, P, L.** *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Judea Pearl. Ser. in Representation & Reasoning. Morgan Kaufmann, 1988, xix + 552 pp, $39.95. [ISBN: 0-934613-73-7] A textbook relating the formal concepts of probability often taught at the advanced undergraduate level to the notion of inference within the context of artificial intelligence. Sections within the text are marked indicating whether they deal primarily with philosophical, technical, or state-of-the-art research concerns. Exercises at the end of each chapter. SM

**Artificial Intelligence, S, P.** *Readings in Nonmonotonic Reasoning.* Ed: Matthew L. Ginsberg. Morgan Kaufmann, 1987, viii + 481 pp, $26.95 (P). [ISBN: 0-934613-45-1] An inference is nonmonotonic if the addition of information may lead to the retraction of an earlier conclusion. This contrasts with conventional deductive inference in which the set of conclusions that can be drawn grows as the set of beliefs grows. This collection of new and previously published papers covers formalisations of nonmonotonic reasoning and applications in artificial intelligence. KS

**Computer Science, P.** *Scientific Computation with Automatic Result Verification.* Ed: U. Kulisch, H.J. Stetter. Comput. Supplementum, V. 6. Springer-Verlag, 1988, viii + 244 pp, $73 (P). [ISBN: 0-387-82063-9] Automatic result verification means that the computer produces, in place of the desired number, two numbers that can be rigorously shown to be upper and lower bounds for the desired number. The papers collected in this volume were for the most part given at the most recent conference (Fall, 1987) at Karlsruhe University where this topic has been a persevering research topic for many years. AWR

**Applications, P, L.** *Distance Geometry and Molecular Conformation.* G.M. Crippen, T.F. Havel. Wiley, 1988, x + 541 pp, $142. [ISBN: 0-471-92061-4] The authors' research on the use of distances and other simple geometric invariants as a means of describing the conformation spaces of the complex organic molecules which occur in biology. In two parts—one on the theoretical approach, and one on algorithmic methods of solving author-developed chemical problems. Claim is that the book is self-contained. Chemical examples in the mathematical portion. Appendices with relevant mathematics. JK

**Applications, L.** *Mathematics Applied to Deterministic Problems in the Natural Sciences.* C.C. Lin, L.A. Segel. Classics in Appl. Math. SIAM, 1988, xxi + 609 pp, (P). [ISBN: 0-89871-229-7] This is an unabridged, corrected republication of the original 1974 Macmillan edition (TR, February 1975; Extended Review, December 1976). It uses a "case study" approach to introduce ideas and methods important in applied mathematics. AO

**Applications, P, L.** *Text, ConText, and HyperText: Writing with and for the Computer.* Ed: Edward Barrett. Ser. in Inform. Syst. MIT Pr, 1988, xxv + 368 pp, $35. [ISBN: 0-262-02275-3] A diverse series of papers about writing documentation, electronic publishing, and professional writing in the context of on-line media and hypertext. Topics range from analysis and poetry to corporate culture, from automated writing to management training. LAS

**Applications, T?(16: 1).** *Non-Life Insurance Mathematics.* Erwin Straub. Springer-Verlag, 1988, 136 pp, $49.80. [ISBN: 0-387-18787-1] Overview of the main actuarial problems and their theoretical solutions in non-life insurance (such as property and liability insurance). Covers such topics as premium calculation, reinsurance, retentions, and reserves. Limited set of exercises. RSK

**Applications (Economics), P.** *Iterative Aggregation Theory.* Lev M. Dudkin, Ilya Rabinovich, Ilya Vakhutinsky. Pure & Appl. Math., V. 111. Marcel Dekker, 1987, xiv + 273 pp, $89.75. [ISBN: 0-8247-7570-8] Iterative aggregation methods can be used to describe hierarchical systems with the problem of coordinating tasks with different levels of information. This book explores these methods in the context of centralised economic planning, with attention given to the more general mathematical theory underlying the methods. SM

**Applications (Economics), P.** *Nonlinear Prefer-*

*ence and Utility Theory.* Peter C. Fishburn. Ser. in Math. Sci., V. 5. Johns Hopkins U Pr, 1988, xiv + 259 pp, $47.50. [ISBN: 0-8018-3598-4] The basic assumptions of expected utility models are often violated by people's reasoned judgments. This book covers developments in preference theory for risky and uncertain decisions. It considers alternate theories of rational preference that take into account systematic departures from expected utility. RH

**Applications (Engineering), S(15-17), L*.** *Higher Mathematics for Engineering Students: Worked Examples and Problems with Elements of Theory, Part 3: Special Courses.* Ed: A.V. Efimov. Transl: Vladimir Shokurov. MIR (US Distr: Imported Pub), 1988, 496 pp, $10.95. [ISBN: 5-03-001131-5] Third in a series of books of worked examples and problems with answers. Probability, statistics, partial differential equations, and integral equations. Excellent and plentiful selection of problems, many non-routine. Each section begins with brief relevant theory and worked examples. Some problems require computer use. In terms of pages per dollar, the book is a terrific buy. JK

**Applications (Engineering), S(15-16), L.** *Worked Examples in Advanced Engineering Mathematics.* L.R. Mustoe. Wiley, 1988, x + 137 pp, $21.95 (P). [ISBN: 0-471-91951-9] Selection of worked examples in linear algebra, eigenvalue problems, optimisation, ordinary and partial differential equations, Fourier series, multivariable calculus, integral transformations, vector fields, complex variables, and statistical methods. From Engineering Council examinations and exam papers set by author and colleagues. Complements conventional textbooks. Brief collection of basic results at book's end; some additional explanation in text. Slim in all aspects but price. JK

**Applications (Fluid Dynamics), P.** *Annual Review of Fluid Mechanics, Volume 21, 1989.* Ed: John L. Lumley, Milton Van Dyke, Helen L. Reed. Annual Reviews, 1989, 479 pp, $34. [ISBN: 0-8243-0721-6] Fourteen survey papers on subjects ranging from turbulent diffusion to boundary-layer separation, and from ocean dynamics to rarefied gases. Cumulative author and title index to all twenty-one volumes. An excellent contribution to a valuable series. LAS

**Applications (Fluid Dynamics), P.** *Dynamics of Curved Fronts.* Ed: Piere Pelcé. Perspect. in Physics. Academic Pr, 1988, xv + 514 pp, $69.50. [ISBN: 0-12-550355-5] Nonlinear dynamics in hydrodynamics, metallurgy, and combustion. Part I is an "introduction to the problem of selection and stability of moving fronts from the flat interface to the complex dendrite." Features unifying concepts. Part II is a compendium of over thirty relevant articles. For researchers and graduate students in chemical and mechanical engineering, physics, and mathematics. JK

**Applications (Physics), S, L.** *Reality and the Physicist: Knowledge, Duration and the Quantum World.* Bernard D'Espagnat. Transl: J.C. Whitehouse, Bernard D'Espagnat. Cambridge U Pr, 1989, 280 pp, $59.50; $19.95 (P). [ISBN: 0-521-32940-X; 0-521-33846-8] The translator has given us a very readable version of this philosophical text by a distinguished physicist. Surprisingly, it is virtually devoid of mathematics. MU

**Applications (Physics), P.** *Nonlinear Stochastic Systems Theory and Applications to Physics.* George Adomian. Math. & Its Applic. Kluwer Academic, 1989, xx + 224 pp, $69. [ISBN: 90-277-2525-X] Describes the use of the decomposition method for the solution of linear and nonlinear stochastic differential equations to solve "real" problems. The emphasis is not on the theory (which is not well-understood), but on how to get numerical answers. AO

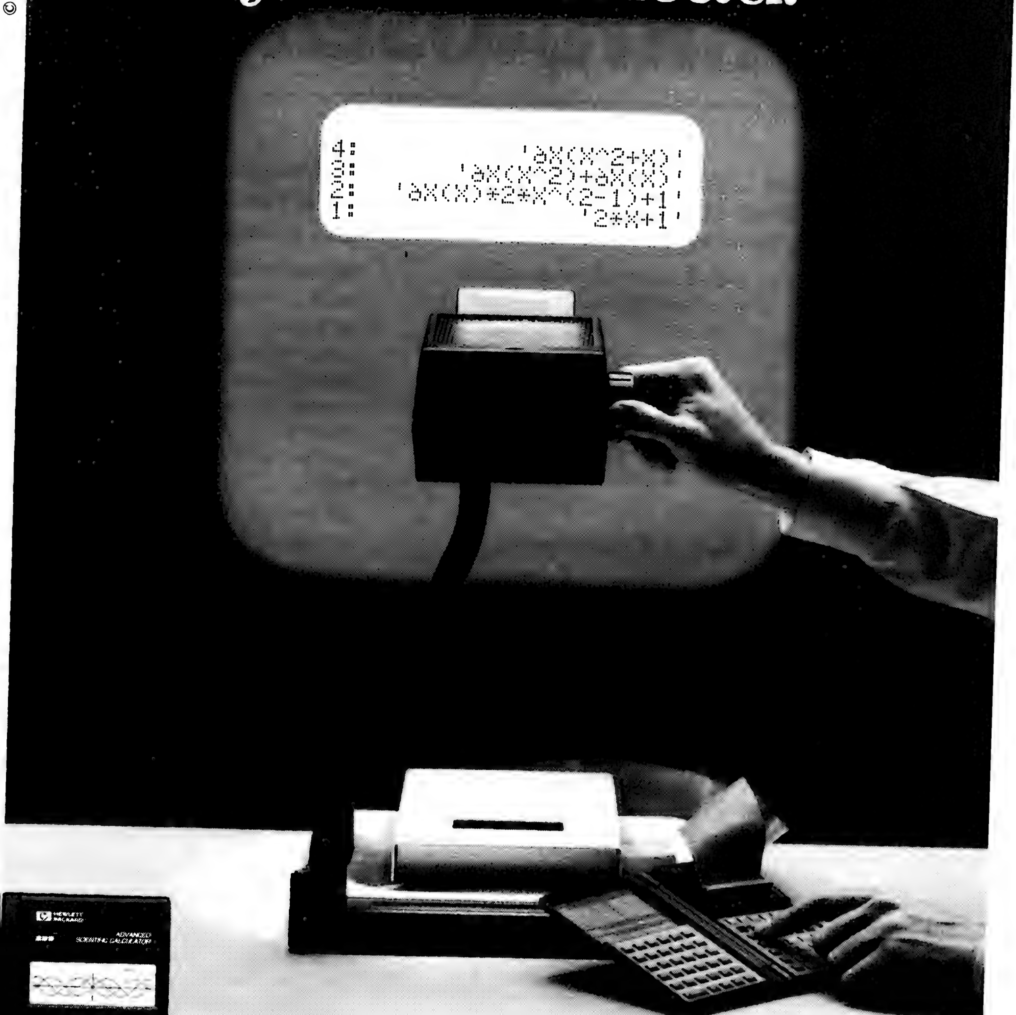
**Applications (Physics), S(18), P.** *Geometry of Classical Fields.* Ernst Bins, Jędrsej Śniatycki, Hans Fischer. Math. Stud., V. 154. North-Holland (US Distr: Elsevier Science), 1988, xviii + 450 pp, $105.25. [ISBN: 0-444-70544-9] Extensive introductory presentation of basic differential geometry, followed by the mathematics of covariant Hamiltonian dynamics. The main part of the book begins with non-relativistic theory, proceeding to Yang-Mills theory and relativistic dynamics. Slightly awkward wording is occasionally distracting. GG

**Applications (Social Science), L.** *Essays on Mathematical Geography.* Sandra Lach Arlinghaus. Institute of Mathematical Geography (IMaGe, 2790 Briarcliff, Ann Arbor, MI 48105), (P). *I*, 1986, Mono. #3, iv + 167 pp, $25.95; *II*, 1987, Mono. #5, v + 101 pp, $12.95. Two parts of a series of typescript publications from "IMaGe," the Institute of Mathematical Geography. These volumes contain a diverse series of papers and reports ranging from the practical (graph theory applied to urban networks; geometry of shadows and solar energy) to the speculative (applying the Heine-Borel Theorem to Middle Eastern politics). Innovative, imaginative, interdisciplinary, and *very* unconventional. LAS

**Reviewers**

RJA: Richard J. Allen, St. Olaf; DFA: David F. Appleyard, Carleton; SB: Steve Benson, St. Olaf; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corsatt, St. Olaf; CE: Christopher Ennis, Carleton; SG: Steven Galovich, Carleton; GG: George Gilbert, St. Olaf; BH: Bruce Hanson, St. Olaf; RH: Rhonda Hatcher, St. Olaf; TH: Timothy Hesterberg, St. Olaf; JJ: Jason Jones, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; GN: Gail Nelson, Carleton; AO: Arnold Ostebee, St. Olaf; SP: Samuel Patterson, Carleton; MR: Matthew Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MS: Myriam Steinback, Macalester; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; LW: Lynne Walling, St. Olaf; PZ: Paul Zorn, St. Olaf.

# Raise the level of teaching by raising your overhead.




Now there's a better way to teach algebra and calculus in the classroom. In fact, two ways.

First, introduce your students to the HP-28S. It's the only calculator that offers symbolic algebra and calculus.

Then, introduce yourself to the overhead display for the HP-28S. It allows you to project your calculations on an HP-28S for everyone in the classroom to see.

**A scholastic offer for you.**

If your department or students purchase a total of 30 HP-28S calculators, we'll give you a classroom overhead display for the HP-28S absolutely free. (A $500 retail value.) Plus, your very own HP-28S calculator free. (A $235 retail value.)

To learn more, and get free curriculum materials, call (503) 757-2004 between 8am and 3pm, PT. Offer ends October 31, 1989.

There is a better way.

HEWLETT PACKARD

Calculator Support, Hewlett-Packard, 1000 NE Circle Blvd., Corvallis, OR 97330

## A History of Algebraic and Differential Topology 1900–1960

**Jean Dieudonné**

This work traces the history of algebraic topology beginning with its creation by Henri Poincaré in 1900 and describing in detail the important ideas introduced in the theory before 1960. Written by a world-renowned mathematician, this book will make exciting reading for anyone working in topology.

1989 / Hardcover / 625 pages / 0-8176-3388-X / $79.00

## Textual Studies in Ancient and Medieval Geometry

**Wilbur R. Knorr**

An important new study of the problems of documentation in ancient technical texts. It presents a survey of the primary Greek evidence on cube duplication together with a selection of medieval Arabic texts on cube duplication and angle trisection, comparing the various versions and describing the changes which occurred as the works were transmitted from one generation to the next. Also an in-depth look at Archimedes' *Dimension of the Circle.*

June 1989 / Hardcover / approx. 640 pages / 0-8176-3387-1 / $89.00

## The Non-Euclidean Revolution

**Richard Trudeau**

In a lively, entertaining and highly accessible style, Trudeau discusses the non-Euclidean Revolution—a scientific revolution every bit as significant as the Copernican revolution in astronomy and the Darwinian revolution in biology. The reader is introduced to the theory of Hyperbolic geometry, an alternative to the traditional Euclidean geometry, which is as logically consistent as Euclid's, and has as much claim to being true as Euclid's, yet extensively contradicts Euclid's.

1987 / Hardcover / 268 pages / 0-8176-3311-1 / $39.00

## Norbert Wiener 1894–1964

**Pesi R. Masani**

This biography provides an incisive look into the life, work and times of one of the great American mathematicians of this Century. Starting with Wiener's childhood fascination for electricity and his early training in mathematical logic with Russell, it traces his profound understanding of the commonality of the functioning of the mind and the computer—to his conception of the subject of *cybernetics*—to his penetrating insight into the impending age of automation in all its ramifications. Wiener's contributions to philosophy, science and engineering are documented and stress is laid on his considerable interaction with other great minds of the times.

June 1989 / Hardcover / approx. 340 pages / 0-8176-2246-2 / $56.50 (tent.)

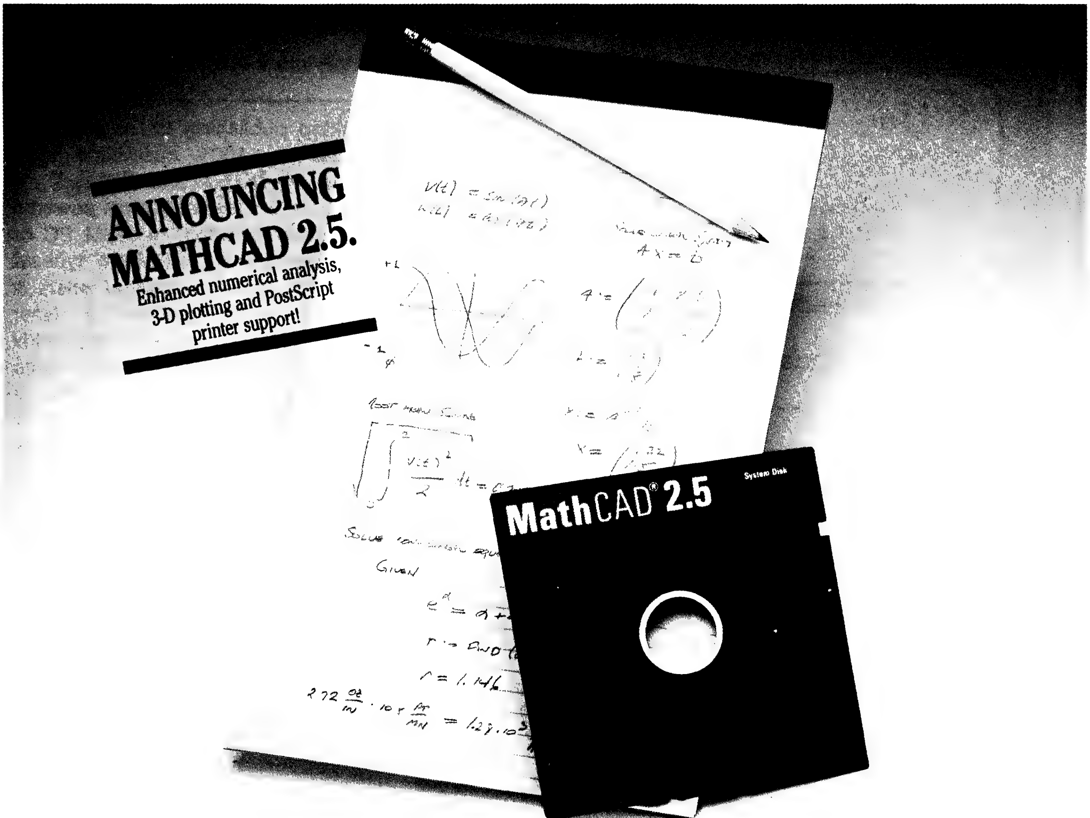
## Newton to Aristotle Towards a Theory of Models for Living Systems

**Mathematical Modeling Series, Vol. 4**

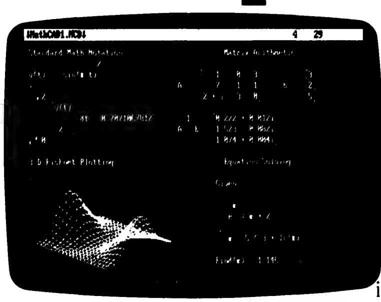
**John Casti and Anders Karlqvist, eds.**

Newton rejected Aristotle's causal theory of scientific events and replaced it with the idea of forces and mechanisms acting upon a collection of material particles. **Newton to Aristotle** examines the manner in which "Aristotelian causation" can be used to fill the modeling gaps left by the Newtonian framework. Using the tools of modern systems theory and mathematics, several internationally renowned scientists demonstrate the advantages of reintroducing the Aristotelian worldview from the perspective of economics, biology, linguistics and other areas of the life and behavioral sciences.

May 1989 / Hardcover / approx. 250 pages / 0-8176-3435-5 / $45.00 (tent.)




# Your pad or ours?



If you perform calculations, the answer is obvious.

MathCAD 2.5.

It's everything you appreciate about working on a scratchpad – simple, free-form math – and more. More power. More accuracy. More flexibility.

Just define your variables and enter your formulas anywhere on the screen. MathCAD formats your equations as they're typed. Instantly calculates the results. And displays them exactly as you're used to seeing them – in real math notation, as numbers, tables or graphs.

MathCAD is more than an equation solver. Like a scratchpad, it allows you to add text anywhere to support your work, and see and record every step. You can try an unlimited number of what-ifs. And print your entire calculation as an integrated document that anyone can understand.

Plus, MathCAD is loaded with powerful built-in features. In addition to the usual trigonometric and exponential functions, it includes built-in statistical functions, cubic splines, Fourier transforms, and more. It also handles complex numbers and unit conversions in a completely transparent way.

Yet, MathCAD is so easy to learn, you'll be using its full power an hour after you begin.

But don't take our word for it – just ask the experts.

PC Magazine recently gave MathCAD their Editor's Choice Award. They described it as "everything you have ever dreamed of in a mathematical toolbox." And when compared to the competition, it was "MathCAD by a mile."

But we didn't stop there. MathCAD 2.5 is a dramatically enhanced version of MathCAD 2.0. MathCAD 2.5 has improved numerical analysis, three-dimensional plotting, and HPGL file import from most popular CAD programs, including AutoCAD.® And for Macintosh® users, there's a MathCAD version 2.0 just for you.

See MathCAD at your local software dealer, or call MathSoft. For information, or a free demo disk, call **1-800-MATHCAD** (in MA, 617-577-1017). Buy MathCAD 2.0 between 5/1/89 and 6/16/89, and get a FREE version 2.5 upgrade (regular upgrade cost is $99 through 6/30/89, and $149 thereafter).

Requires IBM PC® or compatible, 512KB RAM, graphics card.

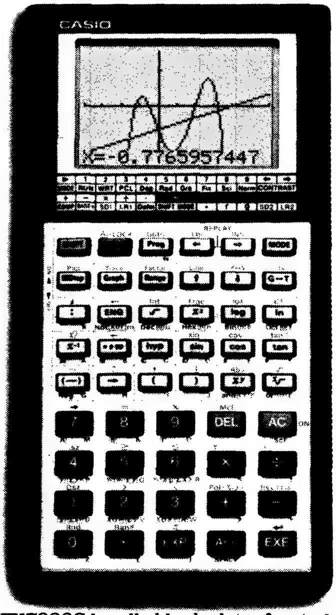
IBM PC® International Business Machines Corporation
MathCAD® MathSoft, Inc

**MathCAD®**

**MathSoft, Inc., One Kendall Sq., Cambridge, MA 02139**

# Casio makes it easy for teachers to make points.

## Introducing the FX7000G/OHP graphic projectable calculator.




*FX7000G handheld calculator for students*

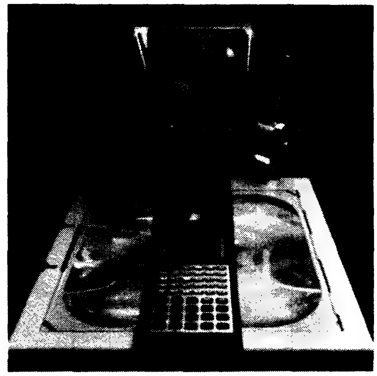
Now it's easy to get your point across. With the Casio FX7000G/OHP graphic projectable calculator.

This new teaching tool combines the FX7000G scientific calculator with an overhead projector capability. This enables all the students in your class to see your work. And to follow along on their own calculators.

The FX7000G is the world's first scientific to convert equations to graphs. At a touch of a button, it computes and graphs complex equations from algebra. Geometry. Calculus. Trigonometry. Physics. Statistics. Even fractals.

It overwrites graphs. Traces points. Magnifies and reduces. Programs and stores individual computations. And features Casio's largest display screen—16 characters by 8 lines.

This powerful scientific assists both teachers and students. In fact, the National Council of Teachers of Mathematics endorses the use of graphing calculators in classrooms across the country.



*Teacher's overhead projectable calculator*

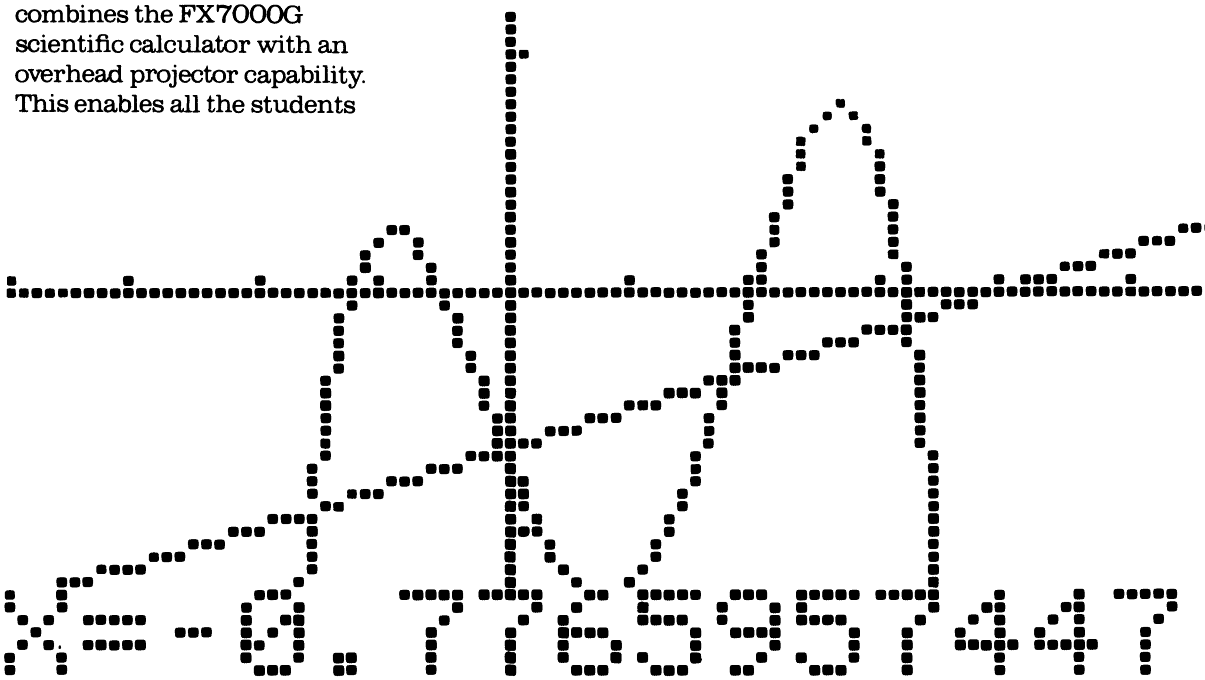
The Casio FX7000G/OHP graphic projectable calculator. Now it's easy to make points with your students.

**For more information, call 1 (800) 553-3338.**

Casio, Inc. Calculator Products Division, 570 Mt. Pleasant Avenue, Dover, NJ 07801.

NCTM MEMBER

CASIO®
Where Miracles Never Cease



THE AMERICAN
MATHEMATICAL MONTHLY

Volume 96, Number 7 August-September 1989

# Contents

(ISSN 0002-9890)

ARTICLES

A Problem of Leo Moser About Repeated Distances on the Sphere . . . . . . . . . . PAUL ERDÖS, DEAN HICKERSON, AND JÁNOS PACH 569

Some Recent Geometric Inequalities . . . . . . . . . . . . . FARUK F. ABI-KHUZAM AND ARTIN B. BOGHOSSIAN 576

Kathy O'Hara's Constructive Proof of the Unimodality of the Gaussian Polynomials . . . . . . . . . . . . . . . . . . . . . . . DORON ZEILBERGER 590

LETTERS TO THE EDITOR . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 602

NOTES

A Cantor Set of Nonconvergence . . . . . . . . . DAVID R. ARTERBURN AND WILLIAM DEAN STONE 604

Merlin's Magic Square Revisited . . . . . . . . . . . . . . . . . . . . . . DANIEL L. STOCK 608

The Number of Words of Length *n* in a Graph Monoid . . . . . . . DAVID C. FISHER 610

On the Number of Furthest Neighbour Pairs in a Point Set . . . . . . . . . HERBERT EDELSBRUNNER AND STEVEN SKIENA 614

A Remark in Divided Differences . . . . . . . . . . . . . . . . . . . . . . . . . . E. T. Y. LEE 618

THE TEACHING OF MATHEMATICS

On $(\mathbf{a}\times\mathbf{b})\times\mathbf{c}$ . . . . . . . . . . . . . . . . . . . . . RONALD SHAW AND FRED J. YEADON 623

The Cauchy-Schwarz Inequality: A Geometric Proof . . . . . . . JAMES W. CANNON 630

From Experimentation to Proof . . . . . . . . . . . . . . . . . . . . . . . . HERVÉ LEHNING 631

Estimating the Diffusion of Stock Prices with the HP-28S . . . YVES NIEVERGELT 636

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 641

Advanced Problems and Solutions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 652

REVIEWS

The Shape of Space, by Jeffrey R. Weeks . . . . . . . . . . . . . . . . . ALAN H. DURFEE 660

Native American Mathematics, edited by Michael P. Closs . . . . JAMES V. RAUFF 662

The Book of Prime Number Records, by Paulo Ribenboim . . CARL POMERANCE 663

TELEGRAPHIC REVIEWS . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 666

## NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to ANITA E. SOLOW, Department of Mathematics, Grinnell College, Grinnell, IA 50112; in geometry or topology to DENNIS DETURCK, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Three copies of proposed problems or solutions should be sent to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, University of Calgary, Alberta, Canada T2N 1N4.




Reprint Permission: MARCIA P. SWARD, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: MS. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $26 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $29. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $90 per year.





Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Assocation of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

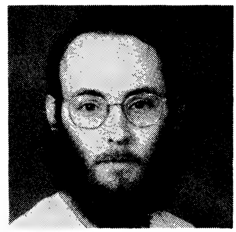
PRINTED IN THE UNITED STATES OF AMERICA

# A Problem of Leo Moser About Repeated Distances on the Sphere

PAUL ERDÖS, DEAN HICKERSON, AND JÁNOS PACH

PAUL ERDÖS: I was born on March 26, 1913. My parents were both mathematics teachers and I learned a great deal of elementary mathematics from them. I received my Ph.D. in 1934 with Professor L. Fejér. I was in Manchester with Professor Mordell from 1934–38 and from 1938 to 1954 I was in the United States. Since then I have constantly travelled around the world: Hungary, the Anglo-Saxon countries, Israel, and Europe.

I have written about 1200 papers and have more than 250 coauthors. My principal subjects are number theory, set theory, combinatorics, geometry, probability, and some branches of analysis.



DEAN HICKERSON: I received my B.S. in Math from the University of California at Davis in 1973, and my Ph.D. from the University of California at Berkeley in 1980. Since then I have worked as a proofreader of mathematical texts and have authored, with Tony Barcellos, the solutions manuals for the calculus textbooks by George Simmons and Sherman Stein. My main research interests are in partitions and in algebraic and combinatorial aspects of geometric tiling problems.

JÁNOS PACH: I was born in Budapest, Hungary, in 1954. I received my Ph.D. from Eötvös University in 1980 under the supervision of Professor M. Simonovits. Since 1977 I have worked at the Mathematical Institute of the Hungarian Academy of Sciences, and held visiting positions at University College London (1981–82), McGill University, Montreal (1984), SUNY at Stony Brook, (1985–86), and Courant Institute, NYU (since 1986). My main fields of research are combinatorics, convexity, discrete and computational geometry.

*Abstract.* We disprove a conjecture of Leo Moser by showing that (i) for every natural number $n$ and $0 < \alpha < 2$ there is a system of $n$ points on the unit sphere $S^2$ such that the number of pairs at distance $\alpha$ from each other is at least const $\cdot\, n \log^* n$ (where $\log^*$ stands for the iterated logarithm function) (ii) for every $n$ there is a system of $n$ points on $S^2$ such that the number of pairs at distance $\sqrt{2}$ from each other is at least const $\cdot\, n^{4/3}$. We also construct a set of $n$ points in the plane in general position (no 3 on a line, no 4 on a circle) such that they determine fewer than const $\cdot\, n^{\log 3/\log 2}$ distinct distances, which settles a problem of Erdös.

**1. Points in the plane.** In most extremal problems in combinatorial and discrete geometry the configurations, arrangements, packings, coverings, etc. which are expected or proved to be optimal, are symmetric in one sense or another. In fact, it is a major obstacle in the way of the research in this field that very few symmetric patterns using a large number of objects are known. Perhaps this is one of the reasons why "latticelike" configurations have attracted so much attention in recent years, and two leading geometers devoted the last couple of years to writing a monograph about "Tilings and Patterns" [GSh]. In spite of the fact that applica-

tions in crystallography and coding theory also inspired extensive computer searches for symmetric configurations, the lack of constructions still remains a general characteristic of the field.

Under these circumstances it is no surprise that so little is known about one of the oldest and most important unsolved problems in discrete geometry: Whether or not the regular lattice packing is the densest packing of equal balls in 3-dimensional Euclidean space. In one of his papers C. A. Rogers made the ironic remark that "many mathematicians believe, and all physicists know" that the answer to this question is affirmative **[R]**. The "knowledge" of the physicists originates in their belief (so often emphasized by Einstein) that the laws of Nature must be simple, and if there existed more economical configurations, then Nature would surely have "invented" them.

In this article we will be concerned with variants of the following questions:

1. What is the minimum number of distances which a set of $n$ points can determine?
2. How many times can a given distance $\alpha$ occur among $n$ points?

We first consider these questions for sets of points in the plane. Beliefs similar to those of the physicists mentioned above led the senior author, more than 40 years ago, to state the following conjectures **[E1]**:

(i) Every set of $n$ points in the plane determines at least $c_1 n/\sqrt{\log n}$ distinct distances (for some constant $c_1 > 0$);
(ii) The number of times a given distance can occur among $n$ points in the plane is at most $n^{1+c_2/\log\log n}$ (for some $c_2 > 0$).

Both bounds are attained for the point system

$$\{(x, y)\colon 0 \leqslant x, y < \sqrt{n},\ x \text{ and } y \text{ are integers}\},$$

i.e., for a $\sqrt{n}$ by $\sqrt{n}$ piece of the integer grid, one of the few known truly symmetric configurations in the plane. In the past many serious attempts were made to attack these problems (see, e.g. **[M1]**, **[Ch]**, **[JSz]**, **[BS]**, **[SSzT]**, **[ChSzT]**, **[EGS]** or the surveys **[E2]**, **[EP]**, **[MP]**), but the gaps between the existing lower and upper bounds are still enormous. (Erdös offered 500 dollars for a proof or disproof of (i) or (ii) several times.)

Due to the small number of known instances of regular point systems in the plane, fighting against these problems is a little bit like shadow boxing: You do not know exactly where the enemy is. The known results in this field reflect the strength (and limits) of the weaponry of combinatorics rather than throw any light on the geometric structure behind. On the other hand, for similar reasons, we must admit that beyond the belief there is very little real evidence supporting the above conjectures.

Even less is known about question 1 under the restriction that the points are in *general position*, i.e. there are no 3 of them on a straight line and no 4 on a circle. We shall need some notation.

Given a set $P = \{p_1, p_2, \ldots, p_n\}$ of $n$ distinct points and a positive number $\alpha$, let

$$\begin{aligned} f(P,\alpha) &= \#\text{ pairs } (p_i, p_j),\ i<j, \text{ at distance } \alpha \text{ from each other},\\ g(P) &= \#\text{ distinct distances determined by pairs of points of } P. \end{aligned} \tag{1}$$

Using this notation, let $G(n) = \min g(P)$, where the minimum is taken over all $n$-element point sets $P$ in the plane in general position. Erdös has asked many times (see, e.g. **[E3]**) the following questions: Is it true that

$$\lim_{n\to\infty} \frac{G(n)}{n} = \infty, \tag{a}$$

$$\lim_{n\to\infty} \frac{G(n)}{n^2} = 0? \tag{b}$$

(Our ignorance in this area is really shocking!) Szemerédi **[Sz]** observed that $G(n) \geqslant (n-1)/3$. (In fact, he conjectures $G(n) \geqslant (n-1)/2$, which would generalize a theorem of Altman **[A]**). Our next result answers question (b) in the affirmative.

THEOREM 1. *For every natural number $n$, $G(n) < (3/2)n^{\log 3/\log 2} < (3/2)n^{1.585}$*

*Proof.* First consider the case $n = 2^k$, and let $P$ be the set of all vertices of the unit cube in $\mathbb{R}^k$, i.e. all $(0,1)$-sequences $x = (x_1, x_2, \dots, x_k)$ of length $k$. Since $x - x'$ is always a $(0, +1, -1)$-sequence, the pairs of distinct points belonging to $P$ determine $3^k - 1$ different vectors. These occur in $(3^k - 1)/2$ pairs of opposite vectors.

One can obviously choose a 2-dimensional plane $\Pi \subseteq \mathbb{R}^k$ such that the orthogonal projection of $P$ onto $\Pi$ is in general position. The projection set $P'$ also determines at most $(3^k - 1)/2$ pairs of opposite vectors, and hence at most this many different distances. Thus $G(2^k) < 3^k/2$.

Now let $n$ be arbitrary. Pick $k$ so that $2^{k-1} < n \leqslant 2^k$. Since $G$ is clearly nondecreasing, we have $G(n) \leqslant G(2^k) < 3^k/2$. But $k < 1 + \log n/\log 2$, so $G(n) < (3/2)3^{\log n/\log 2} = (3/2)n^{\log 3/\log 2}$

Note that the same construction was used in **[DG]** for different purposes. Since the points of $P'$ determine a large number of parallelograms, one cannot resist asking the following question: Does there exist a set $P$ of $n$ points in the plane in general position, such that $P$ does not contain all the vertices of a parallelogram, but $g(P)$, the number of distinct distances determined by $P$, is $o(n^2)$?

**2. Points on the sphere.** What happens if, instead of point systems in the plane, we consider point systems on the sphere? The situation here differs from that in the plane in two important respects. First, there is nothing analogous to the integer lattice, so there are no obvious candidates for the sets which answer questions 1 and 2. Second, the answer to question 2 will depend on the particular distance $\alpha$.

Let $S^{d-1}$ denote the surface of the $d$-dimensional unit ball, i.e.,

$$S^{d-1} = \left\{(x_1, \dots, x_d) : x_1^2 + \cdots + x_d^2 = 1\right\}.$$

More than 20 years ago Leo Moser **[M2]** (see also **[Gu]**, **[MP]**) conjectured that there exists a constant $c$ such that among any $n$ points on the unit sphere $S^2$ the same distance can occur at most $cn$ times, i.e.,

$$f(P, \alpha) \leqslant cn \tag{2}$$

for any $n$-element set $P \subseteq S^2$ and for any $0 < \alpha \leqslant 2$. This conjecture was partly motivated by a well known result conjectured by Vázsonyi and proved independently by several authors (**[G]**, **[H]**, **[St]**), which states that if $P = \{p_1, \dots, p_n\}$ is the

vertex set of a 3-dimensional convex polytope and $\alpha = \max_{1 \leqslant i < j \leqslant n} |p_i - p_j|$, then $f(P, \alpha) \leqslant 2n - 2$.

However, our next theorem shows that Moser's conjecture is false. Let $\log^* n$ denote the minimum integer $r$ such that, starting with $n$, one has to iterate the logarithm function $r$ times to get a value smaller than or equal to 1.

THEOREM 2. *There exist* $c_1, c_2 > 0$ *such that*

(i) *for every natural number* $n$ *and for every* $0 < \alpha < 2$ *one can find* $n$ *points in* $S^2$ *with the property that each is at distance* $\alpha$ *from at least* $c_1 \log^* n$ *others*;

(ii) *for every natural number* $n$ *one can find* $n$ *points in* $S^2$ *with the property that each is at distance* $\sqrt{2}$ *from at least* $c_2 n^{1/3}$ *others.*

*Proof.* (i) Given any $\varepsilon \geqslant 0$, let

$$S_\varepsilon = \{(x, y, z): x^2 + y^2 + z^2 = 1 \text{ and } |z| \leqslant \varepsilon\}.$$

$S_0$ is the *equator* of $S^2$, and $S_\varepsilon$ is called a *strip* of radius $\varepsilon$ around the equator.

Let $0 < \alpha < 2$ be fixed. We also fix a small positive $\varepsilon$ such that

$$2\sqrt{1 - \varepsilon^2} > \alpha, \tag{3}$$

i.e., the diameter of the two circles bounding $S_\varepsilon$ is larger than $\alpha$.

For each $k \geqslant 1$ we shall construct a point set $P$ on the sphere in which each point is at distance $\alpha$ from at least $k$ others. Our construction will be recursive. For $k = 1$, let $P$ consist of 2 points on the equator, at distance $\alpha$ from each other.

To motivate the recursive step, consider the analogous situation in the plane. Given a set $P$ in the plane in which each point is at distance $\alpha$ from at least $k$ others, let $P^* = P \cup \pi(P)$, where $\pi$ is a translation by a vector of length $\alpha$, chosen so that $P \cap \pi(P)$ is empty. Then $P^*$ provides the desired set for $k + 1$. This doesn't work on the sphere because there is no isometry which moves every point the same distance. Instead, we will replace $\pi$ by a set of rotations about a fixed axis, one for each point of $P$.

So suppose that for some $k$ we have a set $P = \{p_1, p_2, \dots, p_{n(k)}\}$ such that each $p_i$ is at distance $\alpha$ from at least $k$ others. Assume further that all points of $P$ are in a narrow strip around the equator, i.e., $P \subseteq S_{\varepsilon(k)}$ for some $\varepsilon(k) < \varepsilon$. Let $u$ and $v$ be two antipodal points on the sphere such that $u$ is at distance $\delta$ from the north pole $(0, 0, 1)$ for some small $\delta$ which will be specified later. We turn $S^2$ around the axis $uv$ so as to bring $p_1$ into a new position $p_1'$ such that $|p_1 - p_1'| = \alpha$. (Note that, in view of (3), this is possible if $\delta$ is sufficiently small.) The rotation of $S^2$ which takes $p_1$ into $p_1'$ and keeps $u$ and $v$ fixed, is denoted by $\pi_1$.

Let $P^{(0)} = P$, $P^{(1)} = \pi_1 P^{(0)}$. If the sets $P^{(0)}, P^{(1)}, \dots, P^{(i-1)}$ have already been determined for some $1 < i \leqslant n(k)$, then we define $P^{(i)}$ as follows. Let $\pi_i$ denote a rotation of $S^2$ around the axis $uv$ for which $|p_i - \pi_i(p_i)| = \alpha$. Set

$$P^{(i)} = \pi_i\left(P^{(0)} \cup P^{(1)} \cup \cdots \cup P^{(i-1)}\right).$$

Finally, let

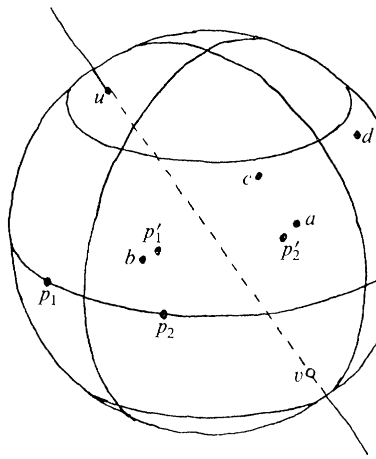
$$P^* = P^{(0)} \cup P^{(1)} \cup \cdots \cup P^{(n(k))}.$$

It is now clear that, for a proper choice of $\delta$ and the axis $uv$, (a) the above definitions are correct, i.e., all $\pi_i$'s exist; (b) the sets $P^{(0)}, P^{(1)}, \dots, P^{(n(k))}$ are pairwise disjoint; (c) there is an $\varepsilon(k + 1) < \varepsilon$ such that $P^* \subseteq S_{\varepsilon(k+1)}$.

Set $n(k+1) = n(k)2^{n(k)}$. According to (b), we have $|P^{(i)}| = 2^{i-1}|P^{(0)}|$ for $1 \leqslant i \leqslant n(k)$, hence $|P^*| = n(k+1)$.

It is now a straightforward matter to show that every point of $P^*$ is at distance $\alpha$ from at least $k+1$ others, which establishes Theorem 2(i) for the numbers $n(k)$. The result then follows easily for all $n$.

The figure below shows this construction for $k = 1$. Here $P^{(1)} = \{p_1', a\}$ is obtained by applying $\pi_1$ to $P^{(0)} = P = \{p_1, p_2\}$. Then $P^{(2)} = \{b, p_2', c, d\}$ is obtained by applying $\pi_2$ to $P^{(0)} \cup P^{(1)}$. The following pairs of points are at distance $\alpha$: $p_1p_2$, $p_1p_1'$, $p_1'a$, $p_2p_2'$, $bp_2'$, $bc$, $cd$, $ad$.




(ii) By a construction due to Erdös (see, e.g., [**Ed** Thm. 6.18]), there exists a positive constant $c_2$ such that one can pick $n/2$ points and $n/2$ lines in the plane, with the property that each of the points lies on at least $c_2n^{1/3}$ of the lines and each of the lines contains at least $c_2n^{1/3}$ of the points. Let $O$ be a point outside the plane supporting this construction.

To each point $P$ of the construction we assign the unit vector pointing from $O$ to $P$. To each line $L$ of the construction we assign one of the two unit vectors perpendicular to the plane determined by $O$ and $L$. This gives $n$ vectors with the property that each of them is perpendicular to at least $c_2n^{1/3}$ others. The endpoints of these vectors lie on the unit sphere centered at $O$ and meet the requirements of (ii).

Note: It follows by the methods used in [**EGS**] that the bound $n^{1/3}$ in Theorem 2(ii) cannot be improved.

Had Moser's conjecture (2) been true, it would have implied that any $n$-element point set $P$ on the sphere $S^2$ determines at least const $\cdot\, n$ different distances. That is, using our notation (1),

$$g(P) \geqslant c'n \tag{4}$$

with an absolute constant $c'$. It is an intriguing open question to decide whether this weaker version of Moser's conjecture is true.

However, it is not hard to show that (4) cannot hold for all $n$-element subsets of any higher dimensional spheres.

THEOREM 3. *For every $d \geqslant 4$, there exists a constant $c_d$ with the property that for infinitely many $n$ one can find an $n$-element point set $P \subseteq S^{d-1}$ determining*

$$g(P) \leqslant \begin{cases} c_4 \dfrac{n}{\log \log n} & \text{if } d = 4 \\ c_d n^{2/(d-2)} & \text{if } d > 4 \end{cases}$$

*different distances.*

We close with some questions suggested by Theorem 2.

Our result for $\sqrt{2}$ is stronger than that for other distances, since $n^{1/3}$ grows faster than $\log^* n$. Is $\sqrt{2}$ really special, or is there a construction which gives a similar result for every $0 < \alpha < 2$? Lacking that, are there such constructions for other particular values of $\alpha$? (Since $\sqrt{2}$ is the edge length of a regular octahedron inscribed in the unit sphere, perhaps the edge lengths of the other Platonic solids are worth investigating.)

We wish to thank Herbert Edelsbrunner for some valuable comments.

REFERENCES

[A] E. Altman, On a problem of P. Erdös, *Amer. Math. Monthly* 70 (1963) 148–157.

[BS] J. Beck and J. Spencer, Unit distances, *J. Combin. Th. A*, 37 (1984) 231–238.

[Ch] F. R. K. Chung, The number of different distances determined by $n$ points in the plane, *J. Combin Th. A*, 36 (1984) 342–354.

[ChSzT] F. R. K. Chung, E. Szemerédi, and W. T. Trotter, Jr., The number of distinct distances determined by a finite point set in the plane (to be published).

[DG] L. Danzer and B. Grünbaum, Über zwei Probleme bezüglich konvexer Körper von P. Erdös und von V. L. Klee, *Math. Zeitschr.*, 79 (1962) 95–99.

[Ed] H. Edelsbrunner, Algorithms in Combinatorial Geometry, Springer-Verlag, Berlin, 1987.

[EGS] H. Edelsbrunner, L. Guibas, and M. Sharir, The complexity of many faces in arrangements of curves and surfaces (to be published).

[E1] P. Erdös, On sets of distances of $n$ points, *Amer. Math. Monthly*, 53 (1946) 248–250. Reprinted in P. Erdös, The Art of Counting, MIT Press, Cambridge, Mass., 1973.

[E2] ______, Problems and results in combinatorial geometry, *Annals of New York Academy of Sciences*, 440 (1985) 1–11.

[E3] ______, On some metric and combinatorial geometric problems, *Discr. Math.*, 60 (1986) 147–153.

[EP] P. Erdös and G. Purdy, Some extremal problems in combinatorial geometry (to appear in *The Handbook of Combinatorics*).

[G] B. Grünbaum, A proof of Vázsonyi's conjecture, *Bull. Res. Council Israel A*, 6 (1956) 77–78.

[GSh] B. Grünbaum and G. Shephard, Tilings and Patterns, W. H. Freeman and Co., New York, 1986.

[Gu] R. K. Guy, Problems, in: The Geometry of Linear and Metric Spaces, Lecture Notes in Math., Vol. 490, Springer, Berlin, 1975, pp. 233–240.

[H] A. Heppes, Beweis einer Vermutung von A. Vázsonyi, *Acta Math. Acad. Sci. Hung.*, 7 (1956) 463–466.

[JSz] S. Józsa and E. Szemerédi, The number of unit distances in the plane. In: Infinite and finite sets, Colloq. Math. Soc. J. Bolyai, Vol. 10, North-Holland, Amsterdam, 1975, 939–950.

[M1] L. Moser, On different distances determined by $n$ points, *Amer. Math. Monthly*, 59 (1952) 85–91.

[M2] ______, Poorly formulated unsolved problems of combinatorial geometry, 1966 (mimeographed).

[MP] W. O. J. Moser and J. Pach, 100 Research Problems in Discrete Geometry, McGill University, Montreal, 1986.

[R] C. A. Rogers, The packing of equal spheres, *Proc. London Math. Soc.* (3) 8 (1958), 609–620.

[SSzT] J. Spencer, E. Szemerédi, and W. T. Trotter, Jr., Unit distances in the Euclidean plane, Graph Theory and Combinatorics: A volume in honour of P. Erdös, Academic Press, London, 1984, 293–303.

[St] S. Straszewicz, Sur un probléme geometrique de P. Erdös, *Bull. Acad. Polon. Sc. CI. III*, 5 (1957) 39–40.

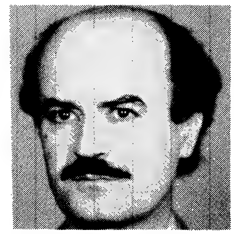
[Sz] E. Szemerédi (unpublished).

# Some Recent Geometric Inequalities

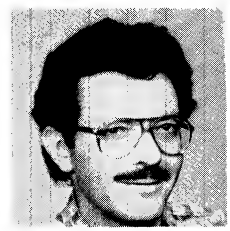
FARUK F. ABI-KHUZAM, *American University of Beirut*

ARTIN B. BOGHOSSIAN, *King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia*

FARUK F. ABI-KHUZAM studied under the supervision of Professor Albert Edrei at Syracuse University where he received his Ph.D. in 1975. He is now Professor of Mathematics at the American University of Beirut, Lebanon. He has held visiting positions at Cornell University (1975) and Syracuse University (1980). In 1987 he was selected a Fulbright Scholar and spent the Summer of 1987 as Honorary Fellow at the University of Wisconsin-Madison.



ARTIN B. BOGHOSSIAN: I received my Ph.D. in 1975 from Syracuse University working in approximation theory under the direction of Professor John Troutman. I returned to my hometown Beirut, Lebanon, as Chairman of the Math-Physics Dept. at Haigazian College (1975–78). After teaching for one year at Yarmouk University, Irbid, Jordan, I moved in 1979 to King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia, where I have since been heavily involved in the academic-administrative functions of the Department of Mathematical Sciences.



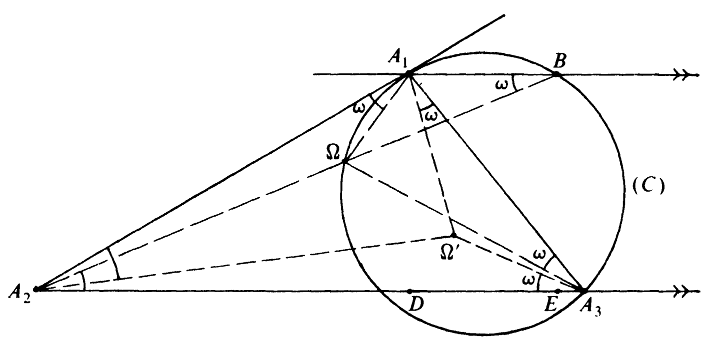
**1. Introduction.** Let $\alpha_1$, $\alpha_2$, and $\alpha_3$ be three *positive* real numbers such that $\alpha_1 + \alpha_2 + \alpha_3 = \pi$, and let $\omega$ be the unique real number in $(0, \pi)$ such that

$$\cot \omega = \cot \alpha_1 + \cot \alpha_2 + \cot \alpha_3. \tag{1}$$

If $\alpha_1$, $\alpha_2$ and $\alpha_3$ are interpreted as the three vertex angles of a triangle $(T) = A_1A_2A_3$, then $\omega$ is called the **Brocard angle** of $(T)$ and there is a point $\Omega$ interior to $(T)$ such that

$$\omega = \angle \Omega A_2 A_3 = \angle \Omega A_3 A_1 = \angle \Omega A_1 A_2.$$

The point $\Omega$ is called a Brocard point of $(T)$ and it may be constructed as follows:



The circle $(C)$ is drawn passing through $A_3$, $A_1$ and tangent to $A_1A_2$ at $A_1$. If the parallel from $A_1$ to $A_2A_3$ meets $(C)$ in $B$, then $A_2B$ intersects $(C)$ in the Brocard point $\Omega$. If perpendiculars $A_1D$ and $BE$ (not shown in the diagram) are dropped onto $A_2A_3$, then one sees immediately that

$$\begin{aligned}\cot\omega &= \overline{A_2E}/\overline{BE} = \overline{A_2E}/\overline{A_1D} = \left(\overline{A_2D} + \overline{DE}\right)/\overline{A_1D}\\ &= \left(\overline{A_2D} + \overline{DA_3} + \overline{A_3E}\right)/\overline{A_1D}\\ &= \cot\alpha_2 + \cot\alpha_3 + \cot\alpha_1,\end{aligned}$$

where one uses the fact (from the diagram) that $\alpha_1 = \angle A_3A_1A_2 = \angle A_3BA_1$ to obtain $\cot\alpha_1 = \overline{A_3E}/\overline{A_1D}$.

If the circle is drawn through $A_1$ and $A_3$ but tangent to $A_3A_2$ at $A_3$, then another point $\Omega'$, also called a Brocard point of $(T)$, may be similarly obtained. But, because of (1), no new Brocard angle is obtained. Thus a triangle has only one Brocard angle $\omega$ given by (1) and two Brocard points $\Omega$ and $\Omega'$. Since

$$\angle\Omega A_1A_2 = \angle\Omega' A_1A_3 = \omega,$$

the lines $\Omega A_1$ and $\Omega' A_1$ (called Cevians) will be symmetric with respect to the bisector of the vertex angle $\alpha_1$. The same is true of the lines $\Omega A_i$ and $\Omega' A_i$ relative to $\alpha_i$ for $i = 2, 3$. This makes $\Omega$ and $\Omega'$ a pair of *isogonal* points, or isogonal conjugates. Other examples of pairs of isogonal points in the triangle are: the circumcenter and the orthocenter; the centroid and the Lemoine point. The incenter is obviously its own isogonal conjugate.

The Brocard points were first noticed in 1816 by Crelle [**6**] but they did not catch the attention of mathematicians of that period for long. In 1875 they were rediscovered by H. Brocard and a considerable amount of work was done on them (see [**8**, p. 263]). The designation $\Omega$ might suggest that they were regarded as the "last" remarkable points to be discovered. But then in 1964, Yff [**17**] discovered a remarkable analogue of the Brocard points and a corresponding **length**, which he called $u$, related to the three sides $a_i$ of a triangle by the equation

$$u^3 = (a_1 - u)(a_2 - u)(a_3 - u). \tag{2}$$

Yff [**17**] proved that

$$8u^3 \le a_1a_2a_3$$

and, on the strength of the analogy between $u$ and $\omega$, conjectured that

$$8\omega^3 \le \alpha_1\alpha_2\alpha_3. \tag{3}$$

Recognized as a gem of elementary geometry, "many mathematicians... tried in vain to prove (3)," as one of them, O. Bottema [**5**] proclaimed. Eleven years were to elapse before a proof of (3) would be found by Abi-Khuzam [**1**]. Describing this proof as "short, elegant and ingenious," Bottema [**5**] was not completely satisfied with the argument leading to the inequality

$$\sin\alpha_1\sin\alpha_2\sin\alpha_3 \le \left[3\sqrt{3}/(2\pi)\right]^3\alpha_1\alpha_2\alpha_3, \tag{4}$$

which was used in the proof of (3). He confirmed the inequality (4) using Lagrange multipliers. Shortly after, using the concavity of $\log(\sin x/x)$, Klamkin [**9**] gave an extension of (4). The original proof in [**1**] of (3) and (4) suggested and made it

possible to establish in [**2**] the inequality

$$\omega^3 \leq (\alpha_1 - \omega)(\alpha_2 - \omega)(\alpha_3 - \omega), \tag{5}$$

which is finer than (3), meaning (5) implies (3). (To see that (5) implies (3), multiply both sides of (5) by $64\omega^3$ and use the algebraic inequality $4\omega(\alpha_i - \omega) \leq \alpha_i^2$.) There followed a series of papers ([**16**], [**10**], [**15**]) repeating the proofs of (3) and (5) but with some variations on the proof of (4), and occasional confusion ([**10**]). For example, the inequality

$$8(\alpha_1 - \omega)(\alpha_2 - \omega)(\alpha_3 - \omega) \leq \alpha_1\alpha_2\alpha_3,$$

which was posed as a problem in [**2**], is stated as fact in [**10**]; it remains, as far as we know, unresolved.

Thus, a certain activity arose around the inequalities (3) and (5), and it was recognized that analysis methods, at least for the present, were indispensable in their treatment. It was a matter of time before efforts would be focused on the problem of improving (3) and (5). A natural approach was to write (3) in the form

$$2\omega \leq \sqrt[3]{\alpha_1\alpha_2\alpha_3},$$

exposing the *geometric mean* of $\alpha_1$, $\alpha_2$, and $\alpha_3$, and to ask whether a similar inequality involving their *harmonic mean,* i.e.,

$$2\omega \leq 3/\left(\alpha_1^{-1} + \alpha_2^{-1} + \alpha_3^{-1}\right), \tag{6}$$

holds true. If valid, (6) would, of course, be a refinement of (3), since the harmonic mean of three positive numbers does not exceed their geometric mean.

This question arose in a different context in works of Stroeker [**14**] and Stroeker-Hoogland [**15**], who posed (6) and

$$3(2\omega)^{-2} \leq \alpha_1^{-2} + \alpha_2^{-2} + \alpha_3^{-2}, \tag{7}$$

along with other inequalities, as conjectures.

A proof of (6) by Mascioni [**11**] appeared at about the same time as submission of this article. A proof of (7), also by Mascioni [**12**], appeared a year later. Both proofs use the method of Lagrange multipliers.

In this article, we shall present proofs of refined versions of (6) and (7) along with some generalizations. We shall also show that all the theory, including identities and inequalities, of the Brocard angle may be derived from the single equation (1). Our main results are summarized below.

THEOREM 1. *Let $\alpha_1$, $\alpha_2$, and $\alpha_3$ be three positive real numbers such that $\alpha_1 + \alpha_2 + \alpha_3 = \pi$, and let $\omega \in (0, \pi)$ be defined by* (1). *Then*

$$\begin{aligned}\omega^{-1} + 3\pi^{-1} &\leq \alpha_1^{-1} + \alpha_2^{-1} + \alpha_3^{-1} \\ &\leq 3(2\omega)^{-1} - (2\cos\omega - \sqrt{3})^2(3\sin 2\omega)^{-1};\end{aligned} \tag{8}$$

$$3(2\omega)^{-2} \leq \alpha_1^{-2} + \alpha_2^{-2} + \alpha_3^{-2} \leq \omega^{-2} - 9\pi^{-2}. \tag{9}$$

*In each of the above four inequalities, equality holds if and only if $\omega = \pi/6$ if and only if $\alpha_1 = \alpha_2 = \alpha_3 = \pi/3$.*

THEOREM 2. *Let $\alpha_1$, $\alpha_2$, $\alpha_3$, and $\omega$ be as in Theorem* 1. *If $\alpha_1$, $\alpha_2$, and $\alpha_3$ are not all equal, then there exists a unique $\beta \in (1,2)$ such that*

$$\alpha_1^{-\lambda} + \alpha_2^{-\lambda} + \alpha_3^{-\lambda} \leq 3(2\omega)^{-\lambda} \quad \text{for all } \lambda \in [0, \beta]; \tag{10}$$

$$\alpha_1^{-\lambda} + \alpha_2^{-\lambda} + \alpha_3^{-\lambda} \geq 3(2\omega)^{-\lambda} \quad \text{for all } \lambda \notin (0, \beta). \tag{11}$$

*In each case, equality holds if and only if $\lambda = 0$ or $\lambda = \beta$.*

Further inequalities may be derived from (11) by first rewriting it in the form

$$\sum_{i=1}^{3} (\omega/\alpha_i)^{\lambda} \geq 3(2^{-\lambda})$$

and, noting that $\omega < \alpha_i$ (to be proved in Section 4), integration with respect to $\lambda$ over the interval $[\lambda, \infty)$, $\lambda \geq \beta$, immediately yields

$$\sum_{i=1}^{3} \alpha_i^{-\lambda} \big/ (\log \alpha_i - \log \omega) \geq 3(2\omega)^{-\lambda} \big/ \log 2 \quad \text{for all } \lambda \geq \beta.$$

Whereas, summation over $\lambda$, $\lambda = 2, 3, 4, \ldots,$ gives

$$\begin{aligned} 3/2 = \sum_{\lambda=2}^{\infty} 3(2^{-\lambda}) &\leq \sum_{\lambda=2}^{\infty} \sum_{i=1}^{3} (\omega/\alpha_i)^{\lambda} \\ &= \sum_{i=1}^{3} . \sum_{\lambda=2}^{\infty} (\omega/\alpha_i)^{\lambda} = \omega^2 \sum_{i=1}^{3} (\alpha_i - \omega)^{-1} \alpha_i^{-1}, \end{aligned}$$

whence we get

$$3(2\omega^2)^{-1} \leq \sum_{i=1}^{3} (\alpha_i - \omega)^{-1} \alpha_i^{-1}. \tag{12}$$

The inequality

$$3\omega^{-1} \leq \sum_{i=1}^{3} (\alpha_i - \omega)^{-1}, \tag{13}$$

which corresponds to (5) and involves the harmonic mean of $(\alpha_1 - \omega)$, $(\alpha_2 - \omega)$ and $(\alpha_3 - \omega)$, was obtained by Abi-Khuzam [**3**]. Now, (13) readily follows from (9):

$$\sum_{i=1}^{3} (\alpha_i - \omega)^{-1} = \omega \sum_{i=1}^{3} (\alpha_i - \omega)^{-1} \omega^{-1} \geq \omega \sum_{i=1}^{3} 4\alpha_i^{-2} \geq 3\omega^{-1};$$

while (12) can also be derived from (8) and (13):

$$\begin{aligned} \sum_{i=1}^{3} (\alpha_i - \omega)^{-1} \alpha_i^{-1} &= \omega^{-1} \sum_{i=1}^{3} \left[ (\alpha_i - \omega)^{-1} - \alpha_i^{-1} \right] \\ &\geq \omega^{-1} \left[ 3\omega^{-1} - 3(2\omega)^{-1} \right] = 3(2\omega^2)^{-1} \end{aligned}$$

It is also possible to obtain reverse inequalities complementing those in (10) and (11). An example of such an inequality for (11) is

$$\alpha_1^{-4} + \alpha_2^{-4} + \alpha_3^{-4} \leq \omega^{-4} - 3^4(13\pi^{-4}), \tag{14}$$

whose proof is indicated in the proof of Theorem 1 in Section 5. However, these questions will not be taken up in this article.

**2. The existence of $\omega$ and preliminary identities.** We now proceed to derive all properties of $\omega$ from (1). Let us then recall that $\alpha_1$, $\alpha_2$, and $\alpha_3$ are *positive* real numbers satisfying $\alpha_1 + \alpha_2 + \alpha_3 = \pi$ and that $\omega$ is the unique real number in $(0, \pi)$ defined by (1).

Using elementary trigonometry and our assumptions on $\alpha_1$, $\alpha_2$, and $\alpha_3$, it is easy to show that

$$\cot \alpha_1 \cot \alpha_2 + \cot \alpha_2 \cot \alpha_3 + \cot \alpha_3 \cot \alpha_1 = 1, \tag{2.1}$$

and

$$\cot \alpha_1 + \cot \alpha_2 + \cot \alpha_3 = \cot \alpha_1 \cot \alpha_2 \cot \alpha_3 + \csc \alpha_1 \csc \alpha_2 \csc \alpha_3. \tag{2.2}$$

Since $\sin \alpha_i > 0$, $i = 1, 2, 3$, and $1 + \cos \alpha_1 \cos \alpha_2 \cos \alpha_3 > 0$, it follows from (2.2) that $\sum_{i=1}^3 \cot \alpha_i > 0$; thus, since $\cot x$ decreases steadily on $(0, \pi)$ and $\cot(\pi/2) = 0$, there exists a unique real number $\omega \in (0, \pi/2)$ such that

$$\cot \omega = \cot \alpha_1 + \cot \alpha_2 + \cot \alpha_3. \tag{2.3}$$

Squaring (2.3) and using (2.1) and $2 \cot \alpha_i \cot \alpha_j \leq \cot^2 \alpha_i + \cot^2 \alpha_j$, we deduce that $\cot^2 \omega \geq 3$, whence

$$0 < \omega \leq \pi/6.$$

Similarly, we readily have

$$\csc^2 \omega = \csc^2 \alpha_1 + \csc^2 \alpha_2 + \csc^2 \alpha_3, \tag{2.4}$$

which is a *differentiated* version of (2.3). To obtain more identities, note that, by (2.1), (2.2), and (2.3), we have the following identity for all $x \in (0, \pi)$:

$$\prod_{i=1}^{3} (\cot x - \cot \alpha_i) = \cot^3 x + \cot x - \cot \omega \csc^2 x + \prod_{i=1}^{3} \csc \alpha_i. \tag{2.5}$$

Putting $x = \omega$ in (2.5), an easy reduction then gives

$$\sin(\alpha_1 - \omega) \sin(\alpha_2 - \omega) \sin(\alpha_3 - \omega) = \sin^3 \omega, \tag{2.6}$$

which is a special case of Ceva's Theorem [**8**, p. 147].

Of course (2.6) is not an *integrated* version of (2.3) but it is reasonably close to being so. It therefore appears plausible to seek identities analogous to (2.3) and (2.4) but involving higher derivatives of the cotangent. Guided by a method of finding the sum of $n$th powers of the roots of an algebraic equation in terms of its coefficients, we may start from (2.5), put $x = \alpha_i$ and sum over $i$ to obtain

$$\sum_{i=1}^{3} \cot^3 \alpha_i + \sum_{i=1}^{3} \cot \alpha_i - \cot \omega \sum_{i=1}^{3} \csc^2 \alpha_i + 3 \prod_{i=1}^{3} \csc \alpha_i = 0. \tag{2.7}$$

Since the derivative of $\csc^2 x$ is $-2 \cot x \csc^2 x$, we use (2.4) to rewrite (2.7) in the

form

$$\begin{aligned}\sum_{i=1}^{3} \cot \alpha_i \csc^2\alpha_i &= \sum_{i=1}^{3} \left(\cot^3\alpha_i + \cot \alpha_i\right) \\ &= \cot \omega \sum_{i=1}^{3} \csc^2\alpha_i - 3\prod_{i=1}^{3} \csc \alpha_i \\ &= \cot \omega \csc^2\omega - 3\prod_{i=1}^{3} \csc \alpha_i\end{aligned} \tag{2.8}$$

and this gives us the identity analogous to (1), but involving the second derivative of $\cot x$.

To obtain the next identity, we multiply (2.5) by $\cot x$, put $x = \alpha_i$ in the resulting equation and sum over $i$. The result is

$$\sum_{i=1}^{3} \cot^4\alpha_i + \sum_{i=1}^{3} \cot^2\alpha_i - \cot \omega \sum_{i=1}^{3} \cot \alpha_i \csc^2\alpha_i + \left(\sum_{i=1}^{3} \cot \alpha_i\right)\prod_{i=1}^{3} \csc \alpha_i = 0. \tag{2.9}$$

If (2.8) and (2.3) are used in (2.9), we arrive at

$$\begin{aligned}\sum_{i=1}^{3} \cot^2\alpha_i \csc^2\alpha_i &= \sum_{i=1}^{3} \left(\cot^4\alpha_i + \cot^2\alpha_i\right) \\ &= \cot^2\omega \csc^2\omega - 4\cot \omega \prod_{i=1}^{3} \csc \alpha_i.\end{aligned} \tag{2.10}$$

The next step is immediate from (2.4) and (2.10):

$$\begin{aligned}\sum_{i=1}^{3} \csc^4\alpha_i &= \sum_{i=1}^{3} \left(\csc^2\alpha_i + \cot^2\alpha_i \csc^2\alpha_i\right) \\ &= \csc^2\omega + \cot^2\omega \csc^2\omega - 4\cot \omega \prod_{i=1}^{3} \csc \alpha_i \\ &= \csc^4\omega - 4\cot \omega \prod_{i=1}^{3} \csc \alpha_i.\end{aligned} \tag{2.11}$$

Noting that the derivative of $\cot x \csc^2 x$ is $-\csc^4 x - 2\csc^2 x \cot^2 x$, we may multiply (2.10) by 2 and add the result to (2.11) to obtain

$$\begin{aligned}&\sum_{i=1}^{3} \left(\csc^4\alpha_i + 2\csc^2\alpha_i \cot \alpha_i\right) \\ &\qquad = \csc^4\omega + 2\csc^2\omega \cot^2\omega - 12\cot \omega \prod_{i=1}^{3} \csc \alpha_i.\end{aligned} \tag{2.12}$$

This is the identity analogous to (1) involving the third derivative of $\cot x$.

It is time now to collect our identities in one compact form that will simplify our proofs later on. As usual, $\cot^{(k)}x$ stands for the $k$th derivative of $\cot x$. Thus, $\cot^{(0)}x = \cot x$, $\cot^{(1)}x = -\csc^2 x$, $\cot^{(2)}x = 2\csc^2 x \cot x$, etc. For the sole purpose of making the form of the next identity simple, we adopt the unusual convention

that $(\cot\omega)^{k-2}$ should be taken to be zero for $k = 0$ or 1. With these conventions, the reader will easily verify that for $k = 0, 1, 2, 3$ and 4, we have:

$$\sum_{i=1}^{3} \cot^{(k)}\alpha_i = \cot^{(k)}\omega - (-1)^k(k+1)!\,(\cot\omega)^{k-2}\prod_{i=1}^{3}\csc\alpha_i. \tag{2.13}$$

Indeed, for $k = 0, 1, 2$ and 3, the formula (2.13) is equivalent to (2.3), (2.4), (2.8), and (2.12), respectively. We trust that the reader will be able to supply a proof of (2.13) for $k = 4$ along the same lines of the previous proofs.

**3. The function $\cot x - x^{-1}$.** We recall that a function $g$ defined on an interval $(a, b)$ is said to be concave if for $0 \le \lambda \le 1$ and $a < x < y < b$ we have

$$\lambda g(x) + (1-\lambda)g(y) \le g[\lambda x + (1-\lambda)y].$$

Using the Mean Value Theorem, it is easy to show that when $g''$ exists on $(a, b)$ then $g$ is concave if and only if $g''(x) \le 0$ there. Furthermore, such a concave $g$ will satisfy the (more general) inequality

$$n^{-1}\sum_{i=1}^{n} g(x_i) \le g\left(n^{-1}\sum_{i=1}^{n} x_i\right)$$

for any set of $n$ points $x_1, x_2, \dots, x_n$ in $(a, b)$.

Our proofs of theorems 1 and 2 are based on the *concavity* and *monotonicity* properties of the function

$$f(x) = \cot x - x^{-1}, \qquad 0 < x < \pi,$$

and its derivatives (Lemma 1 below). The idea is to take one of the identities (2.13) and write it in terms of $f^{(k)}$, the $k$th derivative of $f$. For example, equation (1) is written in terms of $f$ as

$$f(\omega) + \omega^{-1} = \sum_{i=1}^{3}\left[f(\alpha_i) + \alpha_i^{-1}\right]. \tag{3.1}$$

The concavity of $f$ allows us to compare $\Sigma_{i=1}^3 f(\alpha_i)$ and $3f(3^{-1}\Sigma_{i=1}^3\alpha_i) = 3f(\pi/3)$; while the inequality $\omega \le \pi/6$ (see Section 2) allows us to compare $3f(\pi/3)$ with $f(\omega)$ via the monotonicity of $f$. This then reduces (3.1) to an inequality between $\omega^{-1}$ and $\Sigma_{i=1}^3\alpha_i^{-1}$, which is what we want.

The rest of this section is devoted to establishing the concavity and monotonicity of $f^{(k)}$ by showing that $f^{(k)} < 0$ for $k = 0, 1, 2, \dots$. Consider, for example, $f'$ and $f''$ where

$$f'(x) = -\csc^2 x + x^{-2}$$

and

$$f''(x) = 2(\cos x - x^{-3}\sin^3 x)\csc^3 x.$$

Since $\sin x < x$, trivially we have $f' < 0$. But determining the sign of $f''(x)$ requires a comparison of $\cos x$ and $x^{-3}\sin^3 x$ on $(0, \pi)$ which is not a trivial task, as any worker in inequalities would know. (A proof of the inequality $\cos x < x^{-3}\sin^3 x$ on $(0, \pi)$ is given by Mascioni [**11**]; see also Mitrinović [**13**, p. 238].) Moreover, the determination of the sign of the higher derivatives of $f$ by elementary calculus is an even more difficult proposition. To circumvent all these difficulties, we shall treat all

$f^{(k)}$ in a unified manner by utilizing the Euler product representation (e.g., see [4, p. 195]) for the sine function:

$$\sin x = x \prod_{k=1}^{\infty} \left[1 - (x/k\pi)^2\right] \quad \text{for any } x. \tag{3.2}$$

Taking logarithms on both sides of (3.2) and differentiating the resulting equation, we obtain the partial fraction decomposition of $\cot x$:

$$\cot x = x^{-1} + \sum_{k=1}^{\infty} \left[(k\pi + x)^{-1} - (k\pi - x)^{-1}\right], \qquad 0 < x < \pi. \tag{3.3}$$

If each of the functions $(k\pi + x)^{-1}$ and $(k\pi - x)^{-1}$ is expressed as a geometric series in (3.3) and the order of summation is interchanged, we arrive at

$$f(x) = \cot x - x^{-1} = -2x \sum_{k=0}^{\infty} c_k x^{2k}, \qquad 0 < x < \pi, \tag{3.4}$$

where

$$c_k = \sum_{n=1}^{\infty} (n\pi)^{-(2k+2)}, \qquad k = 0, 1, 2, \dots . \tag{3.5}$$

The various operations leading from (3.2) to (3.4) are all justified since the involved series are absolutely and uniformly convergent on closed subintervals of $(0, \pi)$.

Except for $c_0 = \sum_{n=1}^{\infty}(n\pi)^{-2} = 1/6$, we shall not need the exact numerical value of $c_k$. However, (3.5) tells us that $c_k > 0$ for $k = 0, 1, 2, \dots$ . This fact and (3.4) give us the following important lemma.

LEMMA 1. *If $f(x) = \cot x - x^{-1}$ where $0 < x < \pi$, then*

$$f^{(k)}(x) < 0 \quad \textit{for } 0 < x < \pi \quad \textit{and} \quad k = 0, 1, 2, \dots . \tag{3.6}$$

*In particular, the function f and all its derivatives are concave decreasing functions on* $(0, \pi)$.

The function $\log(x^{-1}\sin x)$ being an integral of $f(x)$ is concave by (3.6). There is, however, *another proof* in [1] of the concavity of $\log(x^{-1} \sin x)$, which has the advantage of suggesting other inequalities. This is the proof that was the target of "the slings and arrows of outrageous fortune" in [5], [10], and [16]. It will be taken up in some detail in the next section.

**4. Inequalities for $\alpha_1$, $\alpha_2$, $\alpha_3$, and $\omega$.** If $(i, j, k)$ is a permutation of $(1, 2, 3)$, then our assumption $\alpha_1 + \alpha_2 + \alpha_3 = \pi$, $\alpha_i > 0$, quickly yields

$$\cot \alpha_i + \cot \alpha_j = \sin \alpha_k \csc \alpha_i \csc \alpha_j .$$

Thus, $\cot \alpha_i + \cot \alpha_j > 0$ for $i \neq j$, whence

$$\cot \omega = \cot \alpha_1 + \cot \alpha_2 + \cot \alpha_3 > \cot \alpha_i, \qquad i = 1, 2, 3.$$

It follows that

$$\omega < \alpha_i, \qquad i = 1, 2, 3.$$

(This corresponds to the geometric statement that a Brocard point is interior to its triangle.)

Returning to (2.5) with $x = \omega$, it reduces to

$$\csc\alpha_1 \csc\alpha_2 \csc\alpha_3 = \prod_{i=1}^{3} (\cot\omega - \cot\alpha_i),$$

and we now have that each factor on the right is positive. Applying the inequality between the arithmetic and geometric means of the numbers $\cot\omega - \cot\alpha_i$, $i = 1, 2, 3$, and using (2.3) we obtain

$$\csc\alpha_1 \csc\alpha_2 \csc\alpha_3 \le (8\cot^3\omega)/27.$$

In the opposite direction, starting from the Euler product representation (e.g., see [**4**, p. 195]) for the sine function

$$\sin x = x\prod_{i=1}^{\infty}\left[1 - (x/k\pi)^2\right] \quad \text{for any } x, \tag{4.1}$$

fix an integer $N > 1$ and put

$$P_N(x) = \prod_{k=1}^{N}\left[1 - (x/k\pi)^2\right] = \left[\prod_{k=1}^{N}(1 - x/k\pi)\right]\left[\prod_{k=1}^{N}(1 + x/k\pi)\right].$$

If $0 < \beta_i < \pi$, $i = 1, 2, \ldots, n$, and $\sum_{i=1}^{n}\beta_i = s$, then

$$\prod_{i=1}^{n} P_N(\beta_i) = \left[\prod_{k=1}^{N}\prod_{i=1}^{n}(1 - \beta_i/k\pi)\right]\left[\prod_{k=1}^{N}\prod_{i=1}^{n}(1 + \beta_i/k\pi)\right];$$

utilizing the arithmetic-geometric mean inequalities

$$\prod_{i=1}^{n}(1 - \beta_i/k\pi) \le \left[n^{-1}\sum_{i=1}^{n}(1 - \beta_i/k\pi)\right]^n = (1 - s/nk\pi)^n$$

and

$$\prod_{i=1}^{n}(1 + \beta_i/k\pi) \le \left[n^{-1}\sum_{i=1}^{n}(1 + \beta_i/k\pi)\right]^n = (1 + s/nk\pi)^n,$$

we get

$$\prod_{i=1}^{n} P_N(\beta_i) \le \left[\prod_{k=1}^{N}(1 - s/nk\pi)^n\right]\left[\prod_{k=1}^{N}(1 + s/nk\pi)^n\right] = \left[P_N(s/n)\right]^n.$$

Letting $N \to \infty$ and using (4.1) we arrive at

$$\sin\beta_1 \sin\beta_2 \cdots \sin\beta_n \le \beta_1\beta_2 \cdots \beta_n\left[(n/s)\sin(s/n)\right]^n. \tag{4.2}$$

Taking logarithms on both sides of (4.2) we obtain the concavity of $\log(x^{-1}\sin x)$ alluded to at the end of Section 3.

As an application of (4.2) we have, using Ceva's Theorem (2.6),

$$\begin{aligned}\sin^6\omega &= \sin^3\omega\prod_{i=1}^{3}\sin(\alpha_i - \omega)\\ &\le \omega^3\left[\prod_{i=1}^{3}(\alpha_i - \omega)\right]\left[(6/\pi)\sin(\pi/6)\right]^6.\end{aligned} \tag{4.3}$$

Since $x^{-1}\sin x$ decreases in $(0, \pi/2)$ and $\omega \le \pi/6$, we have

$$(\omega^{-1}\sin\omega)^6 \geqslant [(6/\pi)\sin(\pi/6)]^6.$$

If this inequality is used in (4.3), we obtain (5).

**5. Proof of Theorem 1.** Let $f(x) = \cot x - x^{-1}$ where $0 < x < \pi$, and let $f^{(k)}(x)$ be the $k$th derivative of $f$. By Lemma 1, $f^{(k)}(x)$ is a concave decreasing function on $(0, \pi)$. The concavity of $f^{(k)}$ implies that

$$3^{-1}\sum_{i=1}^{3} f^{(k)}(\alpha_i) \le f^{(k)}\left(3^{-1}\sum_{i=1}^{3}\alpha_i\right) = f^{(k)}(\pi/3), \qquad k = 0,1,2,\dots; \tag{5.1}$$

while, since $f^{(k)}$ is decreasing and $\omega \le \pi/6$, we have

$$f^{(k)}(\pi/6) \le f^{(k)}(\omega), \qquad k = 0,1,2,\dots\,. \tag{5.2}$$

Noting that

$$f^{(k)}(x) = \cot^{(k)}x - (-1)^k k!\, x^{-k-1}, \qquad k = 0,1,2,\dots,$$

and recalling (2.13) with the convention $(\cot\omega)^{k-2} = 0$ for $k = 0$ or 1, and using, (5.1) and (5.2), the following calculations are valid for $k = 0$, 1, 2, 3, and 4.

$$\begin{aligned}
&\cot^{(k)}(\pi/6) - (-1)^k k!\,(6/\pi)^{k+1}\\
&\quad\le \cot^{(k)}\omega - (-1)^k k!\,\omega^{-k-1}\\
&\quad= \sum_{i=1}^{3}\cot^{(k)}\alpha_i + (-1)^k(k+1)!\,(\cot\omega)^{k-2}\prod_{i=1}^{3}\csc\alpha_i - (-1)^k k!\,\omega^{-k-1}\\
&\quad= \sum_{i=1}^{3}\left[f^{(k)}(\alpha_i) + (-1)^k k!\,\alpha_i^{-k-1}\right]\\
&\qquad+(-1)^k(k+1)!\,(\cot\omega)^{k-2}\prod_{i=1}^{3}\csc\alpha_i - (-1)^k k!\,\omega^{-k-1}\\
&\quad\le 3f^{(k)}(\pi/3) + \sum_{i=1}^{3}(-1)^k k!\,\alpha_i^{-k-1}\\
&\qquad+(-1)^k(k+1)!\,(\cot\omega)^{k-2}\prod_{i=1}^{3}\csc\alpha_i - (-1)^k k!\,\omega^{-k-1}.
\end{aligned} \tag{5.3}$$

For $k = 0$ and $k = 1$, (5.3) reduces to

$$\cot(\pi/6) - (6/\pi) \le 3[\cot(\pi/3) - (3/\pi)] + \sum_{i=1}^{3}\alpha_i^{-1} - \omega^{-1}$$

and

$$-\csc^2(\pi/6) + (6/\pi)^2 \leqslant 3\left[-\csc^2(\pi/3) + (3/\pi)^2\right] - \sum_{i=1}^{3}\alpha_i^{-2} + \omega^{-2},$$

respectively, which, in turn, immediately reduce to the first inequality in (8) and the second inequality in (9), respectively. Note that, in each case, if equality holds, then the above steps can be reversed to obtain equality in (5.2) implying $\omega = \pi/6$.

[We note in passing that a proof of (14) may be obtained from (5.3) with $k = 3$ by estimating the term involving $\cot\omega$ and $\csc\alpha_1 \csc\alpha_2 \csc\alpha_3$, using $\cot\omega \geq \sqrt{3}$ and $\sin\alpha_1 \sin\alpha_2 \sin\alpha_3 \leq (\sqrt{3}/2)^3$; this last inequality follows from (4.2).]

To prove the second inequality in (8), let

$$\begin{aligned} g(x) &= f(x) - 2f(x/2) \\ &= \cot x - x^{-1} - 2\left[\cot(x/2) - (x/2)^{-1}\right], \qquad 0 < x < \pi. \end{aligned}$$

It is immediate from (3.4) that $g$ is a concave decreasing function on $(0, \pi)$. So

$$3^{-1} \sum_{i=1}^{3} g(\alpha_i) \leq g\left(3^{-1} \sum_{i=1}^{3} \alpha_i\right) = g(\pi/3) \leq g(2\omega),$$

which implies that

$$3 \sum_{i=1}^{3} \alpha_i^{-1} \leq 9/(2\omega) + 3\cot(2\omega) - 7\cot\omega + 2\sum_{i=1}^{3} \cot(\alpha_i/2), \tag{5.4}$$

and it remains to estimate the last sum in (5.4). From Cauchy's inequality (e.g., see [7, Chapter II]) and (2.4) we obtain

$$\sum_{i=1}^{3} \csc\alpha_i \leq \sqrt{3}\left(\sum_{i=1}^{3} \csc\alpha_i^2\right)^{1/2} = \sqrt{3}\csc\omega, \tag{5.5}$$

which, together with $\cot(x/2) = \cot x + \csc x$ and (2.3), readily yields

$$\sum_{i=1}^{3} \cot(\alpha_i/2) = \cot\omega + \sum_{i=1}^{3} \csc\alpha_i \leq \cot\omega + \sqrt{3}\csc\omega. \tag{5.6}$$

Now, (5.4) and (5.6) give

$$\begin{aligned} 3 \sum_{i=1}^{3} \alpha_i^{-1} &\leq 9/(2\omega) + 3\cot(2\omega) - 5\cot\omega + 2\sqrt{3}\csc\omega \\ &= 9/(2\omega) - (2\cos\omega - \sqrt{3})^2/\sin(2\omega), \end{aligned}$$

and the second inequality in (8) follows. If equality holds, the above steps can be reversed to obtain equality in (5.5) which implies, by invoking the condition of equality in Cauchy's inequality, that $\alpha_1 = \alpha_2 = \alpha_3$ and $\omega = \pi/6$.

To prove the first inequality in (9), we differentiate (3.4) to obtain

$$f'(x) = -\csc^2 x + x^{-2} = -\sum_{k=0}^{\infty} 2(2k+1)c_k x^{2k}, \qquad 0 < x < \pi. \tag{5.7}$$

If we use $c_0 = 1/6$ and let

$$h(x) = 4\left[\csc^2(x/2) - (x/2)^{-2} - 3^{-1}\right] - (\csc^2 x - x^{-2} - 3^{-1}),$$

then (5.7) implies that $h$ is a concave decreasing function on $(0, \pi)$. Thus,

$$3^{-1} \sum_{i=1}^{3} h(\alpha_i) \leq h\left(3^{-1} \sum_{i=1}^{3} \alpha_i\right) = h(\pi/3) \leq h(2\omega).$$

It follows that

$$45(2\omega)^{-2} \le 15\sum_{i=1}^{3}\alpha_i^{-2} + 13\csc^2\omega - 3\csc^2(2\omega) - 4\sum_{i=1}^{3}\csc^2(\alpha_i/2). \quad (5.8)$$

A sharp bound for the last sum in (5.8) is obtained as follows: starting from the identity $4\csc^2 x = \sec^2(x/2) + \csc^2(x/2)$ and using (2.4), we get

$$\sum_{i=1}^{3}\csc^2(\alpha_i/2) = 4\csc^2\omega - 3 - \sum_{i=1}^{3}\tan^2(\alpha_i/2).$$

Since $\sum_{i=1}^{3}\alpha_i = \pi$, trigonometry yields

$$\tan(\alpha_1/2)\tan(\alpha_2/2) + \tan(\alpha_2/2)\tan(\alpha_3/2) + \tan(\alpha_3/2)\tan(\alpha_1/2) = 1.$$

Thus

$$\sum_{i=1}^{3}\tan^2(\alpha_i/2) = \left[\sum_{i=1}^{3}\tan(\alpha_i/2)\right]^2 - 2,$$

and using $\tan(x/2) = \csc x - \cot x$, we have:

$$\begin{aligned}\sum_{i=1}^{3}\csc^2(\alpha_i/2) &= 4\csc^2\omega - 1 - \left[\sum_{i=1}^{3}\tan(\alpha_i/2)\right]^2\\ &= 4\csc^2\omega - 1 - \left[\sum_{i=1}^{3}(\csc\alpha_i - \cot\alpha_i)\right]^2\\ &= 4\csc^2\omega - 1 - \left(\sum_{i=1}^{3}\csc\alpha_i - \cot\omega\right)^2\\ &\ge 4\csc^2\omega - 1 - (\sqrt{3}\csc\omega - \cot\omega)^2\\ &= 2\sqrt{3}\csc\omega\cot\omega,\end{aligned} \quad (5.9)$$

where again we used (5.5). Now, using (5.9) in (5.8) we obtain

$$\begin{aligned}45(2\omega)^{-2} &\le 15\sum_{i=1}^{3}\alpha_i^{-2} + 13\csc^2\omega - 3\csc^2(2\omega) - 8\sqrt{3}\csc\omega\cot\omega\\ &= 15\sum_{i=1}^{3}\alpha_i^{-2} - (8\cos\omega + \sqrt{3})(2\sqrt{3}\cos\omega - 1)(2\cos\omega - \sqrt{3})/\sin^2(2\omega),\end{aligned}$$

and since $\cos\omega \ge \sqrt{3}/2$, the first inequality in (9) follows, with $\omega = \pi/6$ if equality holds.

Finally, if $\omega = \pi/6$, then $\cot\omega = \sqrt{3}$ and

$$\left(\sum_{i=1}^{3}\cot\alpha_i\right)^2 = 3\sum_{i=1}^{3}\cot^2\alpha_i.$$

This implies, by invoking the condition of equality in Cauchy's inequality, that $\alpha_1 = \alpha_2 = \alpha_3 = \pi/3$. Since the remaining converses are trivial, the proof of Theorem 1 is now complete.

**6. Proof of Theorem 2.** If $\lambda < 0$, then $-\lambda > 0$ and, since the numbers $\alpha_i$ are not all equal, the strict inequality between the arithmetic and geometric means of the numbers $\alpha_i^{-\lambda}$ gives

$$\alpha_1^{-\lambda} + \alpha_2^{-\lambda} + \alpha_3^{-\lambda} > 3\sqrt[3]{(\alpha_1\alpha_2\alpha_3)^{-\lambda}} \geq 3(2\omega)^{-\lambda},$$

where we have used (3) to obtain the last inequality. This proves one part of (11).

Strict inequalities hold in (8) and (9) because the numbers $\alpha_i$ are not all equal. Thus, for fixed $\alpha_1$, $\alpha_2$, and $\alpha_3$, the function $g$ defined by

$$g(\lambda) = \sum_{i=1}^{3} (2\omega/\alpha_i)^{\lambda}$$

is a continuous function of $\lambda$ with $g(1) < 3$ and $g(2) > 3$. Therefore, there exists a number $\beta \in (1,2)$ such that $g(\beta) = 3$. Fix one such $\beta$ and consider $\lambda > \beta$. Let $t = (\lambda - \beta)/(\lambda - 1)$ and note that $t \in (0,1)$ and $\beta = t + \lambda(1 - t)$. By Hölder's inequality (e.g., see [7, Chapter II]), we have

$$\begin{aligned} 3(2\omega)^{-\beta} = \sum_{i=1}^{3} \alpha_i^{-\beta} &= \sum_{i=1}^{3} \alpha_i^{-t}\alpha_i^{-\lambda(1-t)} \\ &\leq \left(\sum_{i=1}^{3} \alpha_i^{-1}\right)^{t} \left(\sum_{i=1}^{3} \alpha_i^{-\lambda}\right)^{1-t}, \end{aligned}$$

and using (8) we obtain

$$3(2\omega)^{-\beta} < \left[3(2\omega)^{-1}\right]^{t} \left(\sum_{i=1}^{3} \alpha_i^{-\lambda}\right)^{1-t}.$$

Since $\beta = t + \lambda(1 - t)$, we readily obtain

$$3(2\omega)^{-\lambda} < \sum_{i=1}^{3} \alpha_i^{-\lambda},$$

completing the proof of (11) together with the case of equality.

Finally, suppose that $0 < \lambda < \beta$. Then Hölder's strict inequality yields

$$\sum_{i=1}^{3} \alpha_i^{-\lambda} < 3^{1/p}\left(\sum_{i=1}^{3} \alpha_i^{-\lambda q}\right)^{1/q}$$

for any pair $(p, q)$ of positive numbers satisfying $p^{-1} + q^{-1} = 1$. Since $\lambda < \beta$, we may take $q = \beta/\lambda$. This choice gives

$$\begin{aligned} \sum_{i=1}^{3} \alpha_i^{-\lambda} < 3^{1/p}\left(\sum_{i=1}^{3} \alpha_i^{-\beta}\right)^{1/q} &= 3^{1/p}\left[3(2\omega)^{-\beta}\right]^{1/q} \\ &= 3(2\omega)^{-\beta/q} = 3(2\omega)^{-\lambda}, \end{aligned}$$

proving (10) along with the case of equality.

The uniqueness of $\beta$ being trivial, the proof of Theorem 2 is now complete.

REFERENCES

1. F. F. Abi-Khuzam, Proof of Yff's conjecture on the Brocard angle of a triangle, *Elem. Math.*, 29 (1974) 141–142.
2. ______, Inequalities of Yff-type in the triangle, *Elem. Math.*, 35 (1980) 80–81.
3. ______, A new geometric inequality, *Elem. Math.*, 43 (1988) 75–78.
4. L. V. Ahlfors, Complex Analysis, 2nd edition, McGraw-Hill, New York, 1966.
5. O. Bottema, On Yff's inequality for the Brocard angle of a triangle, *Elem. Math.*, 31 (1979) 13–14.
6. A. L. Crelle, Ueber einige Eigenschaften des ebenen geradlinigen Dreiecks rücksichtlich dreier durch die Winkelspitzen gezogenen geraden Linien, Berlin, 1816.
7. G. H. Hardy, J. E. Littlewood, and G. Pólya, Inequalities, 2nd edition, Cambridge University Press, Cambridge, 1978. (Reprint of 1952 edition.)
8. R. A. Johnson, Advanced Euclidean Geometry, Dover Publications, New York, 1960. (Formerly titled: Modern Geometry, published by Houghton Mifflin Company in 1929.)
9. M. S. Klamkin, On Yff's inequality for the Brocard angle of a triangle, *Elem. Math.*, 32 (1977) 188.
10. L. Kuipers, Extension of Abi-Khuzam's inequality to more than four angles (Dutch), *Nieuw Tijdschr. Wisk.*, 69 (1981/82) 166–169.
11. V. Mascioni, Zur Abschätzung des Brocardschen Winkels, *Elem. Math.*, 41 (1986) 98–101.
12. ______, Zur Abschätzung des Brocardschen Winkels, II, *Elem. Math.*, 42 (1987) 35–42.
13. D. S. Mitrinović, Analytic Inequalities, Springer-Verlag, Berlin-Heidelberg-New York, 1970.
14. R. J. Stroeker, Problem E2905*, *Amer. Math. Monthly*, 88 (1981) 619–620.
15. R. J. Stroeker and H. J. T. Hoogland, Brocardian geometry revisited or some remarkable inequalities, *Nieuw Arch. Wisk.*, 2 (1984) 281–310.
16. G. R. Veldkamp, Elementary proof of an Yff-type inequality (Dutch), *Nieuw Tijdschr. Wisk.*, 69 (1981) 47–51.
17. P. Yff, An analogue of the Brocard points, *Amer. Math. Monthly*, 70 (1963) 495–501.

# Kathy O'Hara's Constructive Proof of the Unimodality of the Gaussian Polynomials

DORON ZEILBERGER*, *Drexel University*

DORON ZEILBERGER received his Ph.D. degree from the Weizmann Institute of Science in 1976, under the direction of Harry Dym. He has been at Drexel University since 1983. His interest in combinatorics began when he tried to find combinatorial applications for his results on partial difference equations. He then became enamored with combinatorics for its own sake. Since 1978 he has attempted to find a constructive proof of the unimodality of the Gaussian polynomials.

**1. Introduction.** Kathy O'Hara has recently [7], [8] solved a long-standing open problem in combinatorics: to find a *direct combinatorial proof* of the unimodality of the coefficients of the Gaussian polynomials. A unimodal sequence $c_0, c_1, \dots, c_n$ is one that increases up to a point and then decreases: i.e. there is an integer $r$ such that $c_0 \le c_1 \le \cdots \le c_r \geqslant c_{r+1} \ge \cdots \ge c_n$.

Combinatorialists love to prove that counting sequences are unimodal. For instance, the prototype of all unimodal sequences is the sequence of binomial coefficients

$$\binom{n}{0}, \binom{n}{1}, \dots, \binom{n}{n-1}, \binom{n}{n}. \tag{1}$$

Many famous combinatorial sequences are unimodal: the Stirling numbers of each kind, for instance. But how can we prove that a sequence is unimodal?

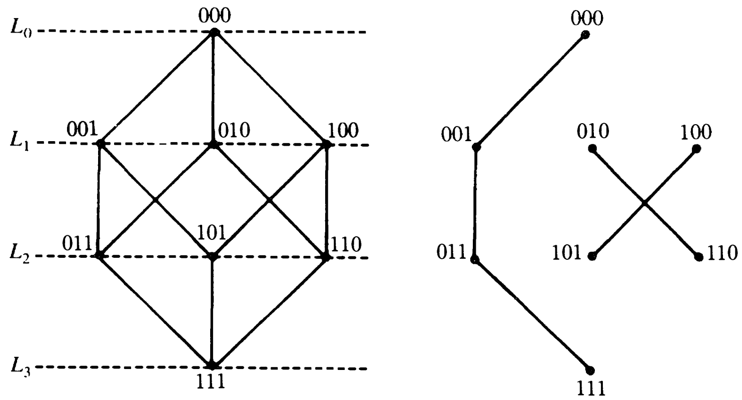
One way is to use analytical methods. If a polynomial has only negative real zeros then its coefficient sequence is unimodal, and even a little more, it is logarithmically convex (see, e.g., [5]). Since the binomial coefficients belong to the polynomial $(1 + x)^n$, which evidently has only negative real zeroes, those coefficients must be unimodal. Since the Stirling numbers of the first kind are the coefficients of the polynomial $(x + 1)(x + 2) \cdots (x + n - 1)$, those numbers must be unimodal too.

Combinatorialists prefer to use combinatorial methods when dealing with such problems, rather than the constructs of analysis. It's often harder, but it's somehow 'purer' too. How might we prove that the binomial coefficients are unimodal in a purely combinatorial way?

Here's one method. Since the coefficients are obviously symmetric about the middle, it will be (necessary and) sufficient to prove that $\left\{\binom{n}{k}\right\}$ increases for $0 \le k \le n/2$. To do that, a fine idea would be to produce, for each fixed $k < n/2$, an injection of the $k$-subsets of $[1, 2, \dots, n]$ into the $k + 1$-subsets. The reader is invited to try to invent such a family of injections, and will then appreciate how hard these proofs can be.

*Supported in part by the National Science Foundation.

Here's another method, that is in fact the one used in O'Hara's proof, the main subject of this paper. Think of the Boolean lattice $B^n$ of all $2^n$ of the subsets of $[1, 2, \dots, n]$, partially ordered by inclusion. Picture the lattice drawn with all $\binom{n}{k}$ of the $k$-subsets laid out on the same horizontal layer, for each $k = 0, 1, 2, \dots, n$, as shown in Fig. 1(a).




(a) $B^3$ (b) An SCD for $B^3$

FIG. 1

Now the binomial coefficients count the number of subsets in each of the layers. Hence what we have to prove is that as we go upstairs in the building, *the layers get fatter and then thinner.*

Well that's a plan, anyway, but how shall we carry it out? By a *chain* we mean a *linearly* ordered collection of these subsets, e.g.,

$$S_1 \subset S_2 \subset \cdots \subset S_r. \tag{2}$$

A chain (2) is *maximal symmetric* if, as we look along the chain from left to right, the cardinalities of the sets increase by exactly 1, and if $\operatorname{card}(S_1) + \operatorname{card}(S_r) = n$.

In Fig. 1(b), maximal symmetric chains are shown as jagged lines that meet every layer from layer $q$ to layer $n - q$, for some $q$.

Suppose we can *cover* the Boolean lattice with pairwise disjoint maximal symmetric chains (such a covering of $B^3$ is shown in Fig. 1(b)). Then we would have proved the unimodality of the coefficient sequence. Indeed, we claim that we would then have a family of injections from the $k$-subsets to the $k + 1$ subsets, for each $k < n/2$. For, given a $k$-subset $S$: map it to the $k + 1$-subset $S'$ that is its successor in the unique maximal chain to which $S$ belongs.

That leaves us with 'merely' the problem of finding a covering of the Boolean lattice with symmetric chains. The reader is invited to think about that one too, and thereby to re-appreciate how hard these proofs can be.

The proof of O'Hara that I want to describe to you is one that results in proving that a certain famous sequence, one that is much more delicate than the binomial coefficients, is unimodal. Hers is the first combinatorial proof of this result, although various non-elementary (i.e., analytic or algebraic) proofs have been

known for some time. The sequence in question (to be described shortly) is touchy: it stays constant for a while, then increases by 1, then stays constant for a while, etc. It is, so to speak, just barely unimodal.

What she did was first to invent a partially ordered set in which the members of the given sequence count the numbers of inhabitants in the layers, and then to invent a maximal symmetric chain decomposition of that partially ordered set. The reader who tries both of our suggested exercises above will appreciate that that was no mean feat.

Specifically, she proved that the polynomial in $q$ ('Gaussian polynomial')

$$G(b,a)=\frac{(1-q^{a+b})(1-q^{a+b-1})\cdots(1-q^{a+1})}{(1-q^{b})(1-q^{b-1})\cdots(1-q)} \tag{3}$$

is unimodal, i.e., its coefficient sequence is unimodal.

For example, $G(2,2)=1+q+2q^2+q^3+q^4$, is clearly unimodal.

The best way to learn a topic is by teaching it. Similarly the best way to understand a new proof is by writing an expository paper about it. That was the original motivation of this paper. Once written, I thought it would be nice if the readers of the Monthly had the opportunity to savor the elegant proof. All the central ideas and constructions are O'Hara's, but I have made a few improvements and shortcuts that I believe will make the argument clearer. I would like to thank Kathy O'Hara for stimulating conversations and correspondence.

The unimodality of the Gaussian polynomials had previously received several 'fancy' proofs, the first one by Sylvester [**15**]. The most elementary proof before O'Hara's was that of Proctor [**10**] who used only linear algebra. The reader is urged to look up Proctor's beautiful paper [**10**] for the history and the significance of the problem. I should also mention White's [**17**] elegant proof that uses Pólya theory.

The coefficients of the Gaussian polynomials $G(b,a)$ have well known combinatorial interpretations, and the polynomials themselves have a number of interesting properties. What follows is a list of some of these. The reader who has not encountered these before could do no better than to consult Pólya, Szegö [**9**], and follow the beautiful exercises 60.1 to 60.11 of Chapter 1, and their solutions, after which the properties listed below will all have been proved.

(a) The coefficient of $q^k$ in $G(b,a)$, let us call it $c_k(b,a)$, is the number of 'zigzag walks' in the plane (i.e., each step increases either $x$ or $y$ by one unit), from $(0,0)$ to $(b,a)$, such that the area under the walk (i.e., between the path, the $x$-axis, and the line $x=a$) is exactly $k$.

(b) $c_k(b,a)$ is the number of '$p$-vectors' in the set

$$U_k(b,a)=\{(p_1,p_2,\dots,p_a)\colon 0\le p_1\le p_2\le\cdots\le p_a\le b,\ p_1+\cdots+p_a=k\} \tag{4}$$

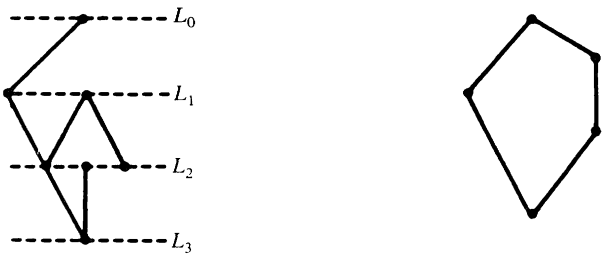
(c) $G(b,a)$ is a polynomial in $q$ of degree $ab$, whose coefficients are symmetric about the middle, i.e., $c_k=c_{ab-k}$.

In order to prove that $G(b,a)$ is unimodal it will be enough to show that $|U_k(b,a)|\le|U_{k+1}(b,a)|$ for $0\le k<ab/2$. A *direct combinatorial proof* will consist of exhibiting an explicit injection of $U_k(b,a)$ to $U_{k+1}(b,a)$ for those values of $k$.

**2. Posets.** We will need some standard definitions and elementary results from the theory of partially ordered sets (*posets*). All that we will need will be presented here, but the reader who wishes to know more is referred to chapter 2 of Stanton and White's excellent text [**14**], and to Greene and Kleitman's survey [**4**].

A poset $(P, \leq)$ is a set $P$ together with a partial order. We say that $a < b$ if $a \leq b$ and $a \neq b$, and that $b$ *covers* $a$ if $a < b$ and there is no $c$ such that $a < c < b$. If $b$ covers $a$ we will write $a \to b$. A poset is *ranked* if each element $a \in P$ has been assigned a positive integer, $\operatorname{rank}(a)$, such that $a \to b$ implies that $\operatorname{rank}(b) = \operatorname{rank}(a) + 1$. The set $L_i = \{a \in P\colon \operatorname{rank}(a) = i\}$ is called the *set of rank* $i$ or the $i$th *level set*.

A good way to visualize a poset $P$ is in terms of its *Hasse diagram*. The vertices of the Hasse diagram are the elements of $P$ and its edges are the covering relations of $P$. The Hasse diagram of a poset uniquely determines that poset, by transitivity of the order relation. Figs. 2(a) and 2(b) show examples of Hasse diagrams of a ranked and an unranked poset respectively.




(a) a ranked poset (b) an unranked poset

FIG. 2

A finite ranked poset is *rank unimodal* if the numbers $w_i = |L_i|$ are unimodal. $P$ is *rank symmetric* if there is an integer $m$ such that $w_i = 0$ for $i > m$ and $w_i = w_{m-i}$ for all $i$. We will call that integer $m$ the rank of $P$.

A *maximal symmetric chain* in a rank symmetric poset is a sequence of elements of $P$, $a_1 \to a_2 \cdots \to a_r$ in which $\operatorname{rank}(a_1) + \operatorname{rank}(a_r) = \operatorname{rank}(P)$. A *maximal symmetric chain decomposition* (SCD) of a rank symmetric poset $P$ is a covering of $P$ by pairwise disjoint maximal symmetric chains.

If a poset has an SCD then it is obviously rank unimodal, and so here is the strategy of the proof.

*Strategy.* Define the poset $U(b, a)$ as follows. Its elements are the $p$-vectors in the set

$$U(b, a) \stackrel{\text{def}}{=} \bigcup_k U_k(b, a)$$
$$= \{(p_1, \ldots, p_a)\colon 0 \leq p_1 \leq \cdots \leq p_a \leq b\}.$$

The partial order relation is simply that $p' < p''$ means that the sum of the entries of $p'$ is less than the sum of the entries of $p''$. We wish to construct an SCD for this poset.

To do this we will need the following three simple facts.

**Fact 1.** If $P$ is a rank-symmetric poset of rank $m$ and if $\bar{P}$ is the same poset but with the rank defined by

$$\overline{\mathrm{rank}}(p) = \mathrm{rank}(p) + \alpha,$$

for some integer $\alpha$, then $\bar{P}$ is a rank-symmetric poset of rank $m + 2\alpha$.

Indeed, $w_i = \bar{w}_{\alpha+i}$ and $w_i = w_{m-i}$ imply that $\bar{w}_{\alpha+i} = \bar{w}_{m-i+\alpha}$.

We say that $\bar{P}$ is the *shift* of $P$ by $\alpha$.

**Fact 2.** Let $P$ and $Q$ be ranked posets and let $\pi$: $P \to Q$ be a rank preserving, order preserving bijection from $P$ to $Q$. If $P$ has an SCD then so does $Q$: its chains are the images, under $\pi$, of the chains of the SCD of $P$.

If $P$ and $Q$ are posets, then their *Cartesian product* $(P \times Q, \leqslant)$ is the poset whose elements are those of the set $P \times Q$, and in which the order relation is defined by $(p, q) \leq (p', q')$ iff $p \leq p'$ and $q \leq q'$. If $P$ and $Q$ are ranked, then $P \times Q$ can be ranked by defining $\mathrm{rank}(p, q) = \mathrm{rank}(p) + \mathrm{rank}(q)$.

**Fact 3.** ([3]) If $P$ and $Q$ are rank-symmetric ranked posets, of ranks $m, m'$ respectively, and if $P$ and $Q$ have SCD's, then $P \times Q$ also has an SCD, and it can be constructed quite explicitly from the two given SCD's. Furthermore, $\mathrm{rank}(P \times Q) = \mathrm{rank}(P) + \mathrm{rank}(Q)$.

The best way to think about the proof of Fact 3 is first to picture the cartesian product of two chains $C$ and $D$ in rectangular form

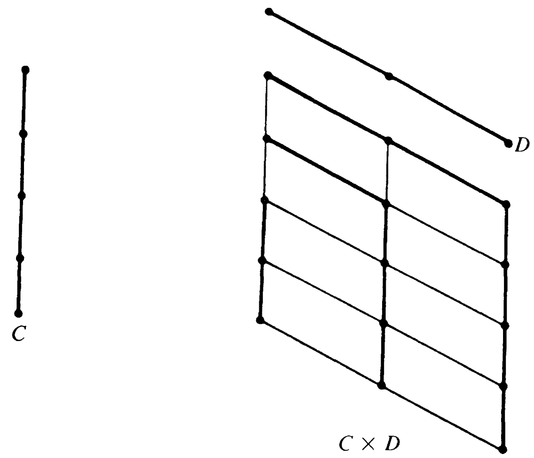
$$\begin{array}{cccc} (p_1, q_1) & (p_2, q_1) & \cdots & (p_r, q_1) \\ (p_1, q_2) & (p_2, q_2) & \cdots & (p_r, q_2) \\ \cdots & \cdots & \cdots & \cdots \\ (p_1, q_s) & (p_2, q_s) & \cdots & (p_r, q_s). \end{array}$$

The chains in the SCD of $P \times Q$ that are contributed by just this one chain $C$ from the SCD of $P$ and the one chain $D$ from the SCD of $Q$ are the ones that are obtained by successively 'peeling off' the chain that is obtained by going from left to right along the top row and continuing all the way down the rightmost column, as shown in Fig. 3.

**3. The art of constructing symmetric chain decompositions.** How does one go about constructing SCD's for a family of posets such as $U(b, a)$? One possibility might have been to express $U(b, a)$ as the cartesian product of smaller $U(b', a')$. For example, for the Boolean lattice $B^n$, we have $B^n = \{0,1\} \times B^{n-1}$, and by applying Fact 3 recursively we obtain an SCD for $B^n$ ([3]). This does not seem possible for $U(b, a)$.

O'Hara's ingenuity consisted in *dividing and conquering*: she found a refinement of $U(b, a)$, a certain doubly indexed family of subposets, $U(b, a; m, d)$, that appeared to be also rank unimodal, rank symmetric, and *of the same rank*, $ab$, as $U(b, a)$. She then discovered a structure theorem for these smaller subposets, that expressed each of them in terms of the operations 'union' and 'Cartesian product' of posets $U(b', a'; m', d')$ for $a' < a$.

It was therefore possible to use the three Facts in order recursively to construct an SCD for each of these $U(b, a; m, d)$ in terms of those of smaller $a$. The base cases $a = -1$, $a = 0$ and $a = 1$ being trivial, this showed that each of the




Decomposition of a product of two chains into chains.

FIG. 3

$U(b, a; m, d)$ had an SCD, and by taking the union over all $m$ and $d$, one had an SCD for $U(b, a)$ itself. We will now present the details.

**4. O'Hara's construction.** Recall that

$$U(b, a) = \{ p = (p_1, \ldots, p_a); 0 \le p_1 \le \cdots \le p_a \le b \}.$$

Let us make the convention henceforth that $p_i = 0$ for $i \le 0$ and $p_i = b$ for $i > a$. For each $p$-vector in $U(b, a)$ we define its *spread* and its *degree*, as follows. First,

$$spread(p) = \max\{ p_i - p_{i-2} \colon 2 \le i \le a + 1 \}.$$

Next we define $M(p)$ to be the set of indices on which the spread is attained,

$$M(p) = \{2 \le i \le a + 1 \colon p_i - p_{i-2} = spread(p)\}.$$

Partition $M(p)$ into $\cup D_j$, where the $D_j$ are maximal *intervals* (of consecutive integers), and define the degree of $p$ by

$$deg(p) = \sum_j \left\lfloor \frac{|D_j| + 1}{2} \right\rfloor.$$

For example, for $a = 13$, $b = 10$ and $p = (1,2,2,3,4,5,6,6,6,7,8,8,10)$, we find that $spread(p) = 2$, $M(p) = \{2,5,6,7,11,13,14\}$, $D_1 = \{2\}$, $D_2 = \{5,6,7\}$, $D_3 = \{11\}$, $D_4 = \{13,14\}$, and $deg(p) = 5$.

Next we define the following subposets:

$$U(b, a; m, d) = \{ p \in U(b, a) \colon spread(p) = m \text{ and } deg(p) = d \},$$

and

$$\bar{U}(b, a; m) = \{ p \in U(b, a) \colon spread(p) \le m \}.$$

We make the reasonable convention that $U(b, 0)$ contains a single $p$-vector, namely the empty vector. If $a$ is negative then $U(b, a) = \varnothing$. Further, we make the

convention that a Cartesian product $\varnothing \times P$, of the empty set with a poset $P$, is isomorphic to $P$, and its elements will be denoted by $(-, p)$.

Everything would follow from

THE O'HARA STRUCTURE THEOREM. *Let $a, b, m, d$ be positive integers. The mapping*

$$\sigma\colon \overline{U}(b - md, a - 2d; m - 1) \times U(ma + 2m - 2b, d) \to U(b, a; m, d),$$

*to be defined below, is an order preserving bijection such that for every $q \in \overline{U}(b - md, a - 2d; m - 1)$ and $r \in U(ma + 2m - 2b, d)$, we have*

$$rank(\sigma(q, r)) = rank(q, r) + 2bd - md(d + 1). \tag{5}$$

Before we go ahead and define $\sigma$, let us show that indeed 'everything' would follow from the theorem.

We prove, by induction on $a = 1, 2, \ldots$, the following proposition: 'For all $b, m, d$, the poset $U(b, a; m, d)$ has an SCD and has rank $ba$.' The base case $a = 1$ being trivial to verify, let $a > 1$ be fixed, and let us assume the truth of the proposition for all $1 \le a' < a$.

We claim first that for every $p \in U(b, a)$ we have $deg(p) \le \lfloor (a + 1)/2 \rfloor$. Indeed, for every $S \subset \{2, \ldots, a + 1\}$, if we write $S = \cup D_j$, where the $D_j$ are intervals, it is easy to check, by induction on $a$, that

$$\sum_j \lfloor (|D_j| + 1)/2 \rfloor \le \lfloor (a + 1)/2 \rfloor,$$

which proves the claim.

Now, if $U(b, a; m, d) \neq \varnothing$ we must have $d \le \lfloor (a + 1)/2 \rfloor$, and since $a > 1$, we have $d < a$. Since $U(ma + 2m - 2b, d)$ is a union of certain $U(ma + 2m - 2b, d; m', d')$, each of which, inductively, has rank $(ma + 2m - 2b)d$ and we know how to construct an SCD for it, it follows that $U(ma + 2m - 2b, d)$ has rank $(ma + 2m - 2b)d$ and we know how to construct an SCD for it. Likewise, $\overline{U}(b - md, a - 2d; m - 1)$ has rank $(b - md)(a - 2d)$ and we know how to construct an SCD for it.

Hence, by Fact 3 we can construct an SCD for the product $\overline{U}(b - md, a - 2d; m - 1) \times U(ma + 2m2b, d)$, and its rank is $(b - md)(a - 2d) + (ma + 2m - 2b)d$. The O'Hara Structure Theorem tells us that $U(b, a; m, d)$ is the image of that product poset under an order preserving bijection, which is actually rank preserving if $U(ma + 2m - 2b, d)$ is shifted by $2bd - md(d + 1)$. By Fact 2 we know how to construct an SCD for $U(b, a; m, d)$, and by Fact 3 its rank is $(b - md)(a - 2d) + (ma + 2m - 2b)d + 2(2bd - md(d + 1)) = ab$.

In order to prove the structure theorem we have to

(i) define the mapping $\sigma$
(ii) define its alleged inverse $\pi$
(iii) prove that $\sigma$ is well defined
(iv) prove that $\pi$ is well defined
(v) prove that the composition $\pi\sigma$ is the identity
(vi) prove that $\sigma\pi$ is the identity
(vii) prove that (5) is true
(viii) prove that $\sigma$ is order preserving.

In order to make this exposition entirely self-contained, we have included complete proofs of these eight propositions in the Appendix. The proof of the unimodality of the Gaussian coefficients is now finished.

**5. Discussion.** The partial ordering that we put on the set of $p$-vectors was the right one to get this proof done, but another partial ordering on the same set of objects has been known for some time, and it is even less well understood. In the set $U(b, a)$ of all $p$-vectors of $\leq a$ components each of which is $\leq b$, a natural order to consider is the one in which $p \leq q$ means that $\forall i$: $p_i \leq q_i$. Under this order $U(b,a)$ is called Young's lattice. More generally, given a partition $\lambda$, the Young lattice $Y_\lambda$ is defined to be the set of all $p$-vectors such that $p \leq \lambda$ with the same definition of $\leq$ and rank as before. The $U(b, a)$ are the special cases in which $\lambda = b^a$ ($b$ repeated $a$ times). $U(b, a)$ is rank unimodal; but what about $Y_\lambda$ for general $\lambda$? Of course $Y_\lambda$ is no longer rank-symmetric, but it appears to be rank-unimodal for small $\lambda$. For example, in $Y_{12}$, the numbers of elements of each rank are $1, 1, 2, 1$.

It was conjectured for some time that the $Y_\lambda$, while not rank-symmetric, are nevertheless rank unimodal. It came as a great surprise when Dennis Stanton [**13**], assisted by a computer, found a counterexample: for $\lambda = (4, 4, 8, 8)$, $Y_\lambda$ is *not* rank unimodal. Furthermore Stanton did something that no computer can do by itself: he found *infinite* families of counterexamples, the simplest one being $(4, 4, 2k, 2k)$, for $k \geq 4$.

A poset that possesses an SCD automatically enjoys the 'Sperner property,' which is to say that its largest possible independent set is obtained by taking the largest level set.

Even the *existence* of an SCD for $U(b, a)$, with Young's lattice order, is still an open problem. However to deduce the Sperner property, as well as rank unimodality, you don't quite need a *symmetric* chain decomposition. It is enough to have a maximal chain decomposition all of whose chains pass through the largest level set. Stanley [**12**], using the 'hard Lefschetz theorem' from algebraic geometry, proved the *existence* of such a chain decomposition for the poset $U(b, a)$ as well as for some other posets, among them the poset $M(n)$ of all partitions of integers into distinct parts $\leq n$, with the same rank and order relation as for $U(b, a)$.

Stanley's proofs for $U(b, a)$ and $M(n)$ were subsequently simplified by Proctor [**10**]. Incidentally, the fact that $M(n)$ has the Sperner property, that Stanley discovered almost as an afterthought, trying to apply the 'hard Lefschetz' hammer to as many nails as possible, turned out to be equivalent to a longstanding conjecture of Erdös and Moser. Stanley found out about it quite by accident, from Larry Harper, during a phone call whose original purpose was to discuss a house sublet during a Sabbatical leave! See Proctor [**10**] for more about this story.

However, both Stanley's proof and Proctor's simplifications were nonconstructive existence proofs. It is still an open problem to find an explicit construction of an SCD for $U(b, a)$ (or even that of a nonsymmetric decomposition as above). To date such a construction is known only when $a$ or $b$ are either 3 or 4 (see Lindstrom [**6**], West [**16**], Riess [**10**]).

I am sure that O'Hara's breakthrough will lead to further work in this area. It would be very nice to find another decomposition of $U(b, a)$ in which the analogue of $\sigma$ would be also order preserving in the Young's lattice order. It would also be

interesting to find an O'Hara-style constructive proof of Stanley's result that the posets $M(n)$ are rank unimodal. This fact is equivalent to the unimodality of the polynomial $(1+t)(1+t^2)\cdots(1+t^n)$.

More generally, Almkvist [1] conjectured that the polynomials

$$\prod_{\nu=1}^{n} \frac{1-t^{r\nu}}{1-t^{\nu}}$$

are unimodal when $r$ is even and $n \geq 1$ and when $r$ is odd and $n \geq 11$. He developed an interesting analytical method that is capable, at least in principle, of proving this conjecture for every *specific* $r$. Assisted by computer, he proved his conjecture for $3 \leq r \leq 20$ and $r = 100, 101$. The conjecture is still open for general $r$. It is equivalent to the rank unimodality of the poset of partitions in which each part can appear at most $r-1$ times, and whose parts are all $\leq n$. An O'Hara-style proof would be particularly gratifying because it would demonstrate that purely bijective methods are capable not only of duplicating results that have been found by other branches of mathematics, but of proving *new* ones.

**Appendix. The fine print (proofs of (i)–(viii)).** It is easiest to start with the definition of $\pi$. So we will perform the eight tasks in the following order: (ii), (i), (iv), (iii), (v), (vi), (vii), (viii).

(*ii*) *Definition of* $\pi$.

$$U(b,a;m,d) \to \overline{U}(b-md, a-2d; m-1) \times U(ma+2m-2b, d)$$

Let $p = (p_1, \dots, p_a) \in U(b, a; m, d)$. We will define $\pi(p) = (q(p), r(p)) = (q, r)$. Let $t+1 = \max M(p) = \max\{2 \leq i \leq a+1;\ p_i - p_{i-2} = m\}$, and define a new $p$-vector $p' = (p'_1, \dots, p'_{a-2})$ by putting $p'_i = p_i$ if $1 \leq i \leq t-2$ and $p'_i = p_{i+2} - m$ else. We will show that $p' \in U(b-m, a-2; m, d-1)$ when $d > 1$ and $p' \in \overline{U}(b-m, a-2; m-1)$ when $d = 1$.

Define $q(p) = p'$, if $d = 1$, and, recursively, $q(p) = q(p')$, if $d > 1$. Also, let $r' = \varnothing$ if $d = 1$, and, recursively, $r' = r(p')$ if $d > 1$ (we will show that $r'$ has $d-1$ nonvanishing components). Finally, put $r_1 = p_t + p_{t+1} + m(a-t-1) - 2b$, and $r_i = r'_{i-1}$ for $2 \leq i \leq d$.

(i) *Definition of* $\sigma$. $\overline{U}(b-md, a-2d; m-1) \times U(ma+2m-2b, d) \to U(b, a; m, d)$

Let $q \in \overline{U}(b-md, a-2d; m-1)$ and $r \in U(ma+2m-2b, d)$; we will define $\sigma(q, r) = p$, say.

Let $r'$ be the $p$-vector with $d-1$ components obtained from $r$ by deleting the first component. Obviously $r' \in U(ma+2m-2b, d-1)$. Next let $p' = q$ if $d = 1$, and $p' = \sigma(q, r')$ if $d > 1$ (we will show that $p'$ has $a-2$ nonvanishing entries).

We will now define a certain integer $t$, $1 \leq t \leq a$. If $r_1 = 0$, let $t = a$. Else, let $t$ be the unique integer that satisfies

$$p'_{t-1} + p'_t - mt < r_1 - m(a+2) + 2b \leq p'_{t-2} + p'_{t-1} - m(t-1). \qquad \text{(A1)}$$

We define $p = \sigma(q, r)$ by

$$p_i = \begin{cases} p'_i, & 1 \le i \le t-1 \\ r_1 - p'_{t-1} - m(a-t+2) + 2b, & i = t \\ p'_{i-2} + m, & t+1 \le i \le a. \end{cases}$$

(iv) $\pi$ is well defined.

$p'$ is a genuine $p$-vector since $p'_{t-1} = p_{t+1} - m = p_{t-1} \ge p_{t-2} = p'_{t-2}$. Here we have used the fact that $t+1 \in M(p)$, and thus $p_{t+1} - m = p_{t-1}$. Now it is readily seen that $M(p') = M(p) \setminus \{t+1\}$ if $t \notin M(p)$, and $M(p') = M(p) \setminus \{t, t+1\}$ if $t \in M(p)$. In either case, $deg(p') = deg(p) - 1$ and $spread(p') = m$, if $d > 1$, while $spread(p') < m$ if $d = 1$.

Hence $p' \in U(b-m, a-2; m, d-1)$ if $d > 1$ and $p' \in \overline{U}(b-m, a-2; m-1)$ if $d = 1$. If $d = 1$, clearly $q \in \overline{U}(b - md, a - 2d; m-1)$, and if $d > 1$, then inductively, $q$ belongs to $\overline{U}(b - m - m(d-1), a - 2 - 2(d-1); m-1) = \overline{U}(b - md, a - 2d; m-1)$. Let $t' + 1 = \max M(p')$. Of course, $t' \le t - 2$. It remains to show that $r$ is a genuine $p$-vector in $U(m(a+2) - 2b, d)$.

By the inductive hypothesis $r'$ belongs to $U(m(a-2+2) - 2(b-m), d-1) = U(m(a+2) - 2b, d-1)$, and so to show that $r$ is a bona fide element of $U(m(a+2) - 2b, d)$ we will have to show that when $d > 1$ we have $0 \le r_1 \le r_2$, i.e., $0 \le r_1 \le r'_1$, i.e., that

$$0 \le p_t + p_{t+1} + m(a-t+1) - 2b \le p'_{t'} + p'_{t'+1} + m((a-2) - t' + 1) - 2(b-m), \tag{A2}$$

and when $d = 1$ we have to show that $r_1 \le m(a+2) - 2b$, i.e.,

$$0 \le p_t + p_{t+1} + m(a-t+1) - 2b \le m(a+2) - 2b.$$

If we make the convention that $t' = -1$ when $d = 1$, then both of the above are included if we prove (A2) for all$d \geqslant 1$. Now the left side of (A2) is equivalent to $(b-m) + b - p_t - p_{t+1} \le m(a-t)$, while the right side of (A2) is equivalent to $p_t + p_{t+1} - p'_{t'} - p'_{t'+1} \le m(t - t')$. Both of these follow from the proposition below: the first by taking $A = a$, $B = t$, and noting that $p_{a+1} = b$ and $p_a \le b - m$; the second by taking $A = t$, $B = t'$, and noting that since $t' \le t - 2$, we have $p'_{t'} = p_{t'}$ and $p'_{t'+1} = p_{t'+1}$.

PROPOSITION. *Let $A \ge B$. Then $p_A + p_{A+1} - p_B - p_{B+1} \le m(A-B)$.*

*Proof of proposition.* If $A - B$ is even then

$$\begin{aligned} p_A + p_{A+1} - p_B - p_{B+1} &= [p_A - p_B] + [p_{A+1} - p_{B+1}] \\ &= [(p_A - p_{A-2}) + \cdots + (p_{B+2} - p_B)] \\ &\quad + [(p_{A+1} - p_{A-1}) + \cdots + (p_{B+3} - p_{B+1}) \\ &\le m(A-B)/2 + m(A-B)/2 = m(A-B). \end{aligned}$$

Similarly, if $A - B$ is odd,

$$\begin{aligned} p_A + p_{A+1} - p_B - p_{B+1} &= [p_A - p_{B+1}] + [p_{A+1} - p_B] \\ &= [(p_A - p_{A-2}) + \cdots + (p_{B+3} - p_{B+1})] \\ &\quad + [(p_{A+1} - p_{A-1}) + \cdots + (p_{B+2} - p_B)] \\ &\le m(A - B - 1)/2 + m(A + 1 - B)/2 = m(A - B). \end{aligned}$$

(iii) $\sigma$ is well defined.

Let $q \in \overline{U}(b - md, a - 2d; m - 1)$ and $r \in U(m(a+2) - 2b, d)$. We will prove that $p = \sigma(q, r)$ is a $p$-vector in $U(b, a; m, d)$.

Since $r' \in U(m(a+2) - 2b, d - 1) = U(ma - 2(b - m), d - 1)$ and $q \in \overline{U}(b - md, a - 2d; m - 1) = \overline{U}((b - m) - m(d - 1), (a - 2) - 2(d - 1); m - 1)$, it follows by the induction hypothesis, for $d > 1$, that $p' \in U(b - m, a - 2; m, d - 1)$. If $d = 1$, then of course $p' \in \overline{U}(b - m, a - 2; m - 1)$.

Now we claim that $p'_{a-1} + p'_a - ma \le r_1 - m(a+2) + 2b \le p'_{-1} + p'_0 - m(1 - 1)$, which, since $p'_{-1} = p'_0 = 0$ and $p'_{a-1} = p'_a = b - m$, is equivalent to $2b - m(a+2) \le r_1 - m(a+2) + 2b \le 0$. That is equivalent to $0 \le r_1 \le m(a+2) - 2b$, which is obvious.

Now it is easy to see that the closed (discrete) interval $[2b - m(a+2), 0]$ can be partitioned as follows into a union of a single point and half open intervals:

$$\begin{aligned} [2b - m(a+2), 0] &= \{2b - m(a+2)\} \\ &\quad \cup \bigcup_{t=a}^{t=1} (p'_{t-1} + p'_t - mt, p'_{t-2} + p'_{t-1} - m(t-1)] \end{aligned}$$

Hence the $t$ that was defined in the definition of $\sigma$ was in fact well defined.

We must also show that $p_t \ge p_{t-1}$ and $p_{t+1} \ge p_t$, both of which follow easily from (A1). Furthermore, $p_a = p'_{a-2} + m \le b - m + m = b$, so $p \in U(b, a)$. Also $t + 1 \in M(p)$.

Now since $r_1 - m(a+2) + 2b \le r_2 - m(a - 2 + 2) + 2(b - m)$, the 'previous $t$,' let us call it $t'$, that was obtained in the previous step out of $r_2$, satisfies $t' \le t - 2$ (if $d = 1$ then $t' = -1$). Thus $M(p) = M(p') \cup \{t + 1\}$ or $M(p) = M(p') \cup \{t, t + 1\}$ (if $d = 1$, $M(p') = \varnothing$). If $d > 1$, then $deg(p) = deg(p') + 1 = d - 1 + 1 = d$. Of course, $spread(p) = m$, and if $d = 1$, $deg(p) = 1$, since $M(p)$ consists of either $\{t\}$ or $\{t, t+1\}$, so in either case $p \in U(b, a; m, d)$.

(vi) $\pi\sigma$ is the identity.

Let $p = \sigma(q, r)$. We have to show that $\pi(p) = (q, r)$. We have just seen that $M(p)$ is obtained from $M(p')$ by adjoining either $t$ alone, or else both $t$ and $t + 1$. In either case $t + 1 = \max M(p)$, so the '$t$ obtained in doing $\pi(p)$' is the same as 'the $t$ obtained in doing $\sigma(q, r)$.' By the inductive hypothesis, if $p' = \sigma(q, r')$ then $\pi(p') = (q, r')$. Finally, 'the $r_1$ obtained from doing $\pi(p)$' is

$$\begin{aligned} & p_t + p_{t+1} + m(a - t + 1) - 2b \\ &\quad = r_1 - p'_{t-1} - m(a - t + 2) + 2b + p'_{t-1} + m + m(a - t + 1) - 2b \\ &\quad = r_1, \end{aligned}$$

as it should.

(v) $\sigma\pi$ is the identity.

Let $(q, r) = \pi(p)$. We have to show that $\sigma(q, r) = p$.

We have $p_t + p_{t+1} = r_1 - m(a - t + 1) + 2b$. But $p_{t+1} = p_{t-1} + m$ since $t + 1 \in M(p)$. So $p_t = r_1 - m(a - t + 2) + 2b - p_{t-1}$. Now $t$ may or may not be in $M(p)$, thus $p_t - p_{t-2} \le m$, which yields the right side of (A2), and $t + 2 \notin M(p)$, so $p_{t+2} - p_t < m$, which yields the left side of (A2). Thus 'the $t$ obtained in doing $\sigma(\pi(p))$' is the same as 'the $t$ obtained in doing $\pi(p)$.' The rest follows from the inductive hypothesis.

(vii) *Proof* that $\operatorname{rank}\sigma(q, r) = \operatorname{rank}(q, r) + 2bd - md - md^2$.

We have

$$\begin{aligned}\operatorname{rank}\sigma(q, r) &= \operatorname{rank}(p) = p_1 + \cdots + p_a\\ &= p_1' + \cdots + p_{t-1}' + \left(r_1 - p_{t-1}' - m(a - t + 2) + 2b\right)\\ &\quad + (p_{t-1}' + m) + \cdots + (p_{a-2}' + m)\\ &= \operatorname{rank}(p') + r_1 + 2(b - m).\end{aligned}$$

First, if $d = 1$ then $p' = q$ and $r = r_1$, so $\operatorname{rank}\sigma(q, r) = \operatorname{rank}(q) + \operatorname{rank}(r) + 2(b - m)$, which was to be shown in this case. Next, if $d > 1$, the inductive hypothesis gives

$$\begin{aligned}\operatorname{rank}\sigma(q, r) &= \operatorname{rank}\sigma(q, r') + r_1 + 2(b - m)\\ &= \operatorname{rank}(q) + \operatorname{rank}(r') + 2(b - m)(d - 1)\\ &\quad - m(d - 1)d + r_1 + 2(b - m)\\ &= \operatorname{rank}(q) + \operatorname{rank}(r) + 2bd - md(d + 1)\\ &= \operatorname{rank}(q, r) + 2bd - md(d + 1).\end{aligned}$$

(viii) $\sigma$ is order preserving.

This is immediate from part (vii) since the order on $U(b, a)$ is the trivial one in which $p < q$ iff $\operatorname{rank}(p) < \operatorname{rank}(q)$.

## REFERENCES


1. Gert Almkvist, Proof of a conjecture about unimodal polynomials, *J. of Number Theory*, to appear.
2. George Andrews, The Theory of Partitions, Addison Wesley, Reading, MA, 1976.
3. N. G. de Bruijn, C. van E. Tenbergen, and D. Kruyswijk, On the set of divisors of a number, *Nieuw. Arch. Wisk.*, (2) 23 (1952) 191–193.
4. Curtis Greene and Daniel Kleitman, Proof techniques in the theory of finite sets, in *Studies in Combinatorics*, G. C. Rota, ed., Mathematical Association of America, 1978, pp. 22–79.
5. G. H. Hardy, J. E. Littlewood, and G. Pólya, Inequalities, Cambridge University Press, 1952.
6. B. Lindstrom, A partition of $L(3, n)$ into saturated chains, *European J. Comb.*, 1 (1981) 60–63.
7. Kathleen M. O'Hara, Unimodality of Gaussian coefficients: a constructive proof, research announcement, preprint, University of Iowa.
8. ______, Unimodality of Gaussian coefficients: a constructive proof, *J. Comb. Theory*, Ser. A, to appear.
9. G. Pólya and G. Szegö, Problems and Theorems in Analysis, Springer-Verlag, New York, 1972.
10. Robert Proctor, Solution of two difficult combinatorial problems with linear algebra, this MONTHLY 89 (1982) 721–734.
11. W. Riess, Zwei optimizierungprobleme auf Ordnungen, *Arbeitsberichte des Instituts für mathematische Maschinen und datenverarbeitungen*, 11 (1978) no. 5 Erlangen.

12. Richard P. Stanley, Weyl groups, the hard Lefschetz theorem, and the Sperner property, *SIAM J. Alg. Discr. Meth.*, 1 (1980) 164–184.
13. Dennis Stanton, Unimodality and Young's lattice, *J. Comb. Th.*, to appear.
14. Dennis Stanton and Dennis White, Constructive Combinatorics, Springer-Verlag, New York, 1986.
15. James J. Sylvester, Proof of the hitherto undemonstrated fundamental theorem of invariants, Collected Math. Papers, vol. 3, Chelsea, New York, 1973, pp. 117–126.
16. Douglas B. West, A symmetric chain decomposition of $L(4, n)$, *European J. Comb.*, 1 (1980) 379–383.
17. Dennis E. White, Monotonicity and unimodality of the pattern inventory, *Adv. Math.*, 38 (1980) 101–108.

# LETTERS TO THE EDITOR

Editor:

While I agree with Joseph Fulda (Material implication revisited, this MONTHLY, March, 1989, 247–250) that there is no need to revise the truth table for the conditional connective of propositional logic in order to explain most uses of the words "if" and "then," or "implies," I would like to suggest an alternate explanation for the widespread confusion on this point. Many sentences containing the words "if" and "then" should properly be interpreted as containing not only a conditional, but also an unstated universal quantifier. The belief that the truth table for the conditional cannot be used to explain the meaning of such sentences is, I believe, a result of failing to recognize the presence of this quantifier, and does *not* indicate any defect in the truth table for the conditional. If the truth table for the conditional is properly combined with the meaning of the quantifier, the resulting interpretation of the sentence is in complete agreement with intuition.

For example, we all know that for sequences of real numbers, monotone and bounded implies convergent, although neither monotone nor bounded alone is sufficient to imply convergent. But according to the conditional truth table, the statements $(M \wedge B) \to C$ and $(M \to C) \vee (B \to C)$ are equivalent. (If you don't believe it, make the truth tables!) Does this mean that the conditional cannot be used to interpret the word "implies" in this case? The problem is not with the conditional, but with a missing quantifier. When we say, "Monotone and bounded implies convergent," what we really mean is, "For all sequences $s$, if $s$ is monotone and bounded then $s$ is convergent." Inserting the missing quantifiers, we find that we should have been comparing the statements $\forall s[(M(s) \wedge B(s)) \to C(s)]$ and $\forall s(M(s) \to C(s)) \vee \forall s(B(s) \to C(s))$. These are not equivalent; in fact, interpreting "$M$," "$B$," and "$C$" as meaning "monotone," "bounded," and "convergent," respectively, on the universe of sequences of real numbers gives a counterexample to the equivalence. I do not see how one can explain the trouble in this example, as Fulda suggests, by claiming that there is a confusion between "if" and "iff." One need not think that all convergent sequences are monotone and bounded to be puzzled by this example.

Even some nonmathematical "paradoxes" involving the conditional can be explained as missing quantifiers. For example, consider the sentence, "If the sun is shining then it must be between 2:00 and 3:00 PM." The conditional truth table

seems to suggest that the sentence is true at night, or on cloudy days, when the sun is not shining. If you think the sentence is false at all times, you are probably thinking of it as a general statement, perhaps more precisely expressed as, "For all times $t$, if the sun is shining at time $t$ then time $t$ is between 2:00 and 3:00 PM." Since there are times which are not between 2:00 and 3:00 PM when the sun is shining, the standard truth table for the conditional can now be used to explain why this sentence is false at all times. Once again, the confusion was caused not by any peculiarity in the conditional truth table, but by the failure to recognize the presence of an unstated quantifier.

DAN VELLEMAN
Amherst College
Amherst, MA 01002

Editor:

The plausibility argument given by Professor Donald Hartig for the differentiation formula for the sine (On the Differentiation Formula for $\sin\theta$, this MONTHLY (March, 1989), 252) is hardly new. In fact, the argument appears (using infinitesimals and equalities rather than approximations) in a treatise of Roger Cotes, *On the Estimation of Errors*, published in 1721, six years after Cotes' death, in an article "De planetarium stationibus" by F. C. Maier, a colleague of Euler in St. Petersburg, in 1727, and in the 1737 text of Thomas Simpson, *A New Treatise on Fluxions*. More details can be found in my article, "The Calculus of the Trigonometric Functions," *Historia Mathematica* 14 (1987), 311–324.

Sincerely,

VICTOR J. KATZ
University of the District of Columbia

# NOTES

EDITED BY DENNIS DETURCK, DAVID J. HALLENBECK, AND ANITA E. SOLOW

## A Cantor Set of Nonconvergence

DAVID R. ARTERBURN and WILLIAM DEAN STONE
*Department of Mathematics, New Mexico Institute of Mining and Technology, Socorro, NM 87801*

**Introduction.** Define a sequence by repeated mappings of a number by a function. There are three "nice" classes of such sequences: sequences that converge to fixed points, those that are (or converge to) periodic sequences, and those that become unbounded or undefined. Frequently these are the only possibilities. It is possible, however, to have bounded sequences that do not converge at all. We examine the set of points that lead to such non-convergent sequences for a class of functions from an interval to itself. We find that Cantor-like sets, those most ubiquitous of pathological sets, arise quite naturally in this problem.

**A Motivating Example.** The January 1986 issue of *The College Mathematics Journal* [1] poses the problem:

For what value of $x_0$ does the sequence $x_{n+1} = 3 - 2/x_n$ converge?

Let $f(x) = 3 - 2/x$ for $x \neq 0$. It is fairly straightforward to show that 1 and 2 are fixed points; $f(0)$ is undefined; there is a sequence of points $0 = y_0 < y_1 < y_2 < \cdots$ such that $f(y_{n+1}) = y_n$. For $x_0$ any other point we get a sequence that converges to 2.

A more interesting problem is obtained by leaving $f$ the same for $x > 0$, then taking the odd extension. That is, consider the sequence $x_{n+1} = f(x_n)$, where

$$f(x) = \left\{\begin{matrix} -3 - 2/x: & x < 0 \\ \text{undefined}; & x = 0 \\ 3 - 2/x; & x > 0 \end{matrix}\right\} = \operatorname{sgn}(x)[3 - 2/|x|].$$

Whereas before the interval $(-\infty, 0)$ was mapped onto $(3, \infty)$, and thus under repeated iterations to 2, we now have sequences $\{f^n(x)\}$, for $x < -1$, converging to $-2$; $x = -1$ is a fixed point. For $x \geqslant 1$ the sequence $\{f^n(x)\}$ is exactly the same as for our original function; for $x > 1$ the sequence converges to 2, 1 is a fixed point. What about sequences of iterates of points in the interval $(-1, 1)$?

THEOREM 1. *The set $S$ of points $x$ in $(-1, 1)$ such that the sequence of iterates $\{f^n(x)\}$ does not converge to $\pm 2$ is a Cantor-like (nowhere dense, perfect) set.*

In fact, it is a Cantor-like set of measure zero, as will be shown below.

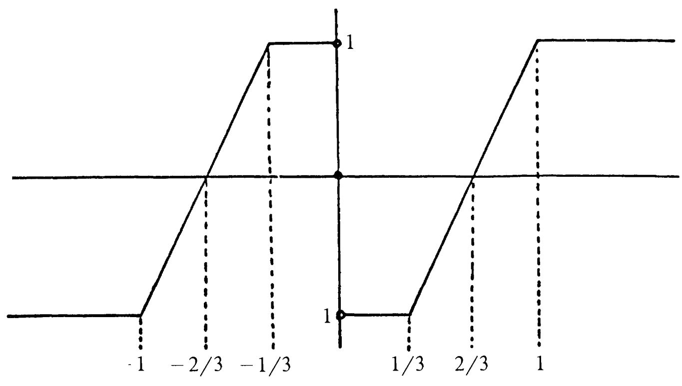
*Proof.* The interval $(-1/2, 0)$ is mapped to $(1, \infty)$ and thence toward 2. $(0, 1/2)$ goes to $(-\infty, -1)$ and thence to $-2$. The map is undefined at 0. $(-1, -1/2)$ and $(1/2, 1)$ are both mapped onto $(-1, 1)$. Mapping again repeats the process, taking the center out of the two remaining intervals. Further iteration constructs a Cantor-like set; a portion out of the middle of each interval remaining after $n$ iterations is mapped to the interval $(-1/2, 1/2)$ and thence either it is mapped to

one of the fixed points or (if the $n$th iteration of $f$, $f^n(x) = 0$) the sequence becomes undefined. Since we have removed an open interval from the interior of each closed interval whose image after $n$ iterations is still in $[0, 1]$, and this open interval does not abut the boundary of the closed interval, we have constructed a Cantor-like set. For each interval in the complement of this Cantor-like set there is a single point in the interior that is mapped (eventually) to 0 and after this further iterations are undefined. The interval left of this point goes to 2, the right to $-2$. We thus get a countable set of points that are mapped, after a finite number of iterations, to 0. Everything else in the complement of the Cantor-like set converges under repeated mapping to 2 or $-2$.

**A Piecewise Linear Example.** What is the behavior of the remaining points, the Cantor-like set?

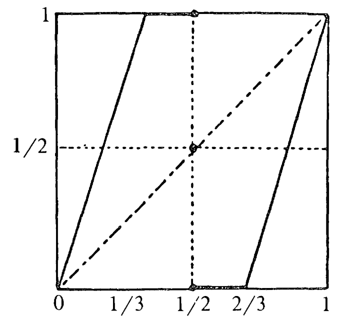
To make this question more accessible let us change our mapping. Since everything outside of the interval $[-1, 1]$ eventually goes to one of the fixed points, we will define $f^*(x) = 1$ for $x \geqslant 1$, $f^*(x) = -1$ for $x \leqslant -1$. We want the middle third (except for 0) mapped outside of $(-1, 1)$ and the outside thirds each mapped onto $(-1, 1)$. A piecewise linear odd function that does all this is this:

$$f^*(x) = \begin{cases} 1; & x \geqslant 1 \text{ or } -1/3 \leqslant x < 0 \\ 3x - 2; & 1/3 \leqslant x \leqslant 1 \\ 0; & x = 0 \\ 3x + 2; & -1 \leqslant x \leqslant -1/3 \\ -1; & 0 < x \leqslant 1/3 \text{ or } x \leqslant -1. \end{cases}$$




By scaling we can change this function to a map of the unit interval into itself:

$$h(x) = \begin{cases} 3x; & 0 \leqslant x \leqslant 1/3 \\ 1; & 1/3 \leqslant x < 1/2 \\ 1/2; & x = 1/2 \\ 0; & 1/2 < x \leqslant 2/3 \\ 3x - 2; & 2/3 \leqslant x \leqslant 1. \end{cases}$$




If we express $x$ in ternary form this works out to be:

$$\text{if } x = \begin{cases} .0abc\ldots \\ \\ .1abc\ldots, \\ \\ .2abc\ldots \end{cases} \quad f(x) = \begin{cases} .abc\ldots \\ 1 \text{ if first non-1 digit is } 0 \\ 1/2 \text{ if all digits are } 1 \\ 0 \text{ if first non-1 digit is } 2 \\ .abc\ldots \end{cases}$$

Thus if the leading digit is 1, $x$ is sent to a fixed pint, otherwise the digits are shifted one to the left with the first digit removed. Any point with a 1 anywhere in its expansion (or a tail of $000\ldots$ or $222\ldots$) is sent to a fixed point after a finite number of iterations. This leaves precisely the Cantor middle thirds set less the (countable) set of all the endpoints of the intervals in the successive approximations. This set is uncountable and consists of all points whose ternary expansions are non-terminating and include no 1's.

What happens to these points? Sequences generated from the rational numbers in the Cantor set eventually become periodic with period equal to the length of the repeating block in the number's ternary expansion.

*Example.*

$$x_0 = .200220220, \quad x_1 = .00220220, \quad x_2 = .022022,$$

$$x_3 = .22022022, \quad x_4 = .2022022, \quad x_5 = x_2, x_6 = x_3, \quad \text{etc.}$$

The irrational elements of the Cantor set generate sequences that do not converge to fixed points or periodic solutions. For any sequence of positive integers $a$, $b$, $c, \ldots$ there exists an $x_0$ such that the sequence $x_0, x_1, x_2, \ldots$ is in $(0, 1/3)$ for $a$ iterations, then in $(2/3, 1)$ for $b$ iterations, then in $(0, 1/3)$ for $c$ iterations, etc.

*Example.* If $\mathrm{a, b, c}\ldots = 1, 1, 2, 2, 3, 3, \ldots$ let $\mathrm{x}_0 = .020022000222\ldots$(in base three).

**Generalization.** The results in this paper hold for an entire family of functions. If the function $f$ is such that there are numbers $a < b < c < d$ such that the intervals $[a, b]$ and $[c, d]$ are both mapped continuously onto $[a, d]$ while the interval $(b, c)$ is mapped (either immediately or asymptotically) to fixed points then the first

theorem holds: the set of points *not* mapped to a fixed point is a Cantor-like set. A counting argument shows that for each integer $n > 2$ there is at least one solution of period $n$ ($f^n$ maps $2^n$ disjoint subintervals continuously onto $[a, d]$).

If $f' > 1$ on $[a, b]$ and $[c, d]$ it can be shown that there are finitely many periodic points of period $n$, and only finitely many points that are mapped to one of these periodic points in any given number of iterations. Thus there are uncountably many non-periodic, non-converging points. The proofs are left as an exercise for the reader. If (as in our motivating example) the derivative of $f$ is greater than 2 a.e., we can also prove:

THEOREM. 2 *Let* $S = \{x | f^n(x)$ *is an element of* $(a, b)$ *or* $(c, d)$ *for all* $n\}$. *Then* $m(S) = 0$.

*Proof.* Define $E$ to be $\{x \text{ in } S | x > 0\}$. Let $g:(a, b) \to (a, d)$ be $f$ restricted to $(a, b)$. $f(S) = S$, with the mapping 2-to-1, and by symmetry $g(E) = S$, with the mapping 1-to-1. Again by symmetry, if $m(S) = \alpha$, $m(E) = \alpha/2$. Now since $g^{-1}(S) = E$ it is known [2] that for $g$ absolutely continuous and monotone increasing,

$$m(S) = \int_E g'(x)\, dx.$$

If $g' > 2$ a.e., as it is in our problem, this gives $m(S) = (2 + h)m(E)$, where $h > 0$. However, $m(E) = m(S)/2$, so we get $m(S) = [(2 + h)/2]m(S)$, thus $m(S) = 0$. Thus, as in our piecewise linear problem we have a Cantor set of measure zero consisting of points which do not converge to a fixed point.

It *is*, however, possible to construct a function such that theorem 2 does not hold. i.e., the Cantor set of nonconverging points has measure greater than zero, if the derivative is less than or equal to two on a set of positive measure. For an example let $S$ be the Cantor set obtained from the interval $[0, 1]$ by removing at the $n$th step a closed interval of length $1/4^n$ from the middle of each interval; $m(S) = 1/2$. We define $f$ to have fixed points at $0, 1/2$ and $1$; to be continuous except at $1/2$; and such that $f(x) = 1$ for $x$ in $[3/8, 1/2)$, $f(x) = 0$ for $x$ in $(1/2, 5/8]$. On each interval $[a, b]$ in $[0, 3/8]$ intersected with the complement of $S$ define $f$ by $f(x) = 4x - (a + b)/4$, that is a line segment with slope 4 and center on the line $y = 8x/3$. For intervals $[c, d]$ in $[5/8, 1]$ intersected with the complement of $S$ define $f(x) = 4x - (c + d)/4 - 5/3$, a line segment of slope 4 and center on the line $y = 8x/3 - 5/3$. Define $f$ on $S$ so that $f$ is continuous (the slope must average 2 on $S$).

With $f$ defined like this $f^{-1}$ (a multivalued function) maps the complement of $S$ onto itself. $f^{-1}(0) = [3/8, 1/2) \cup \{0\}$; $f^{-1}(1) = (1/2, 5/8] \cup \{1\}$; $f^{-1}(1/2) = \{3/16, 1/2, 13/16\}$. These are all in the complement of $S$, so points in $S$ are never mapped to a fixed point. In fact one can show that the measure of the set of points mapped to a fixed point after $n$ iterations will be $1/2\ [1 - (1/2)^{n+1}]$, and $S$ is exactly the set of points never mapped to one of the fixed points.

Further, in a manner similar to Yorke & Li's work [3] we can prove:

THEOREM 3. *For any sequence of positive integers* $\{a_i\}_{i=1}^{\infty}$ *there exists an* $x_0$ *such that the sequence* $\{x_n\}$ *is in* $[a, b]$ *for the first* $a_1$ *terms, then in* $[c, d]$ *for* $a_2$ *terms, in* $[a, b]$ *for* $a_3$, *in* $[c, d]$ *for* $a_4$, *etc.*

*Proof.* Define the sequence of intervals $I_n$ by

$$I_n = \begin{cases} [a, b] & n = 0, 1, \dots, a_1 - 1 \\ [c, d] & n = a_1, a_1 + 1, \dots, a_1 + a_2 - 1 \\ [a, b] & n = a_1 + a_2, \dots, a_1 + a_2 + a_3 - 1 \\ \text{etc.} & \end{cases}$$

In other words the first $a_1$ intervals are $[a, b]$, the next $a_2$ are $[c, d]$, the next $a_3$ are $[a, b]$, and so forth.

Now define $Q_0 = I_o$. Since $f(I_n) = [a, d]$ for all $n$, we have $I_1$ contained in $f(Q_0)$. Thus there is a closed interval $Q_1$ contained in $Q_0$ such that $f(Q_1) = Q_0 = I_0$. If $Q_{n-1}$ has been defined so that $f^{n-1}(Q_{n-1}) = I_{n-1}$, $I_n$ is in $f(I_{n-1}) = f^n(Q_{n-1})$ so there is a $Q_n$ in $Q_{n-1}$ such that $f^n(Q_n) = I_n$. Any $x_0$ in $Q_n$ is such that $f^n(x_0) = x_n$ is in $I_n$. Let $Q$ = the intersection over $n$ of $\{Q_n\}$. Since the $Q_n$'s are a nested set of compact intervals the intersection is not empty. If $x_0$ is in $Q$, then $x_n$ is in $I_n$ for all $n$, and the theorem is proved.

REFERENCES

1. H. J. Fletcher, problem 318, *The College Mathematics Journal* (Jan, 1986) 92.
2. H. L. Royden, Real Analysis, 2nd edition, Macmillan, 1968.
3. T. Y. Li and J. A. Yorke, Period Three Implies Chaos, *Amer. Math. Monthly*, 82 (1975) 985–992.

## Merlin's Magic Square Revisited

DANIEL L. STOCK
*R. R. Software, Inc., 2145 Crooks Road, Troy, MI 48084*

In a recent article [1], Don Pelletier described a game based on a Parker Brothers toy called MERLIN. Although several games can be played on MERLIN, we consider only the one Pelletier examined, called Magic Square. Pelletier gave an algorithm, involving simple calculations on a certain matrix, for solving the game in a minimal number of moves. In this note, we use Pelletier's result to derive a simpler algorithm, allowing the game to be solved in a minimal number of moves by inspection.

The relevant parts of the toy are nine translucent buttons, arranged in a square array. The buttons are numbered from 1 to 9, with the top row labelled 1 to 3 from left to right, the middle row labelled 4 to 6 from left to right, and the bottom row labelled 7 to 9 from left to right. Each button can be in one of two states—OFF or FLASHING—distinguishable by means of a light behind the button. Pushing a button toggles the states of certain nearby buttons, in a fixed, symmetrical way. The game starts with the buttons being some pattern of independent, random states. The goal of the game is to push a sequence of buttons to achieve a set winning pattern of states.

We shall refer to any button as being one of the four *corner* buttons (numbers 1, 3, 7, and 9), the unique *center* button (number 5), or one of the four remaining *edge* buttons (numbers 2, 4, 6, and 8). The winning pattern consists of all the corner and edge buttons flashing, but the center button off.

To finish our description of the game, we only need to specify more precisely the effect of pushing any one button. The following rules apply:

—Pushing an edge button toggles the state of that button and the two adjacent corner buttons;

—Pushing a corner button toggles the state of that button, the two adjacent edge buttons, and the center button; and

—Pushing the center button toggles the state of that button and all four edge buttons.

Pelletier used several important observations in finding his solution. First, note that the order of pushing buttons is insignificant. Also, pushing any one button twice is equivalent to not pushing it at all. Combining these results, we see that for any sequence of buttons to be pushed, there is an equivalent sequence in which no button is pushed more than once. Since order is insignificant, one need only establish a set of buttons to be pushed, rather than a sequence. Pelletier showed that for any given starting pattern, there is a set of buttons that solves the game; furthermore, this set is unique, so that it represents the minimal solution.

Pelletier offered the following algorithm for finding this minimal solution for a given starting pattern. Consider the following matrix $A^{-1}$, which is the binary inverse of $A = [A_{i,j}\colon A_{i,j} = 1$ if pushing button $j$ toggles the state of button $i$, 0 otherwise]:

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

Given an initial pattern $P$, select a subset $S_P$ of the columns of $A^{-1}$ as follows: the fifth column is in $S_P$ iff button 5 is flashing; for $n$ not equal to 5, the $n$th column is in $S_P$ iff button $n$ is off. Let $A_P^{-1}$ denote the submatrix of $A^{-1}$ obtained by retaining only the columns of $A^{-1}$ whose indices belong to $S_P$. Then the set of buttons to push includes button $n$ iff the sum, modulo 2, of the entries in the $n$th row of $A_P^{-1}$ is 1.

The reason that button number 5 is treated oppositely from all the other buttons in Pelletier's algorithm is that it is the center button, the only one that is off in the specified winning pattern. By treating other buttons in the same way, this algorithm can be modified to work for any desired winning pattern.

We can simplify Pelletier's algorithm by exploiting the inherent symmetry in the problem and by imposing an artificial sequencing on the solution. The sequencing allows us to use intermediate information in the toy to derive the minimal solution more simply. A sequencing that works particularly well is to first consider pushing the edge buttons, then the corner buttons, and finally the center button. The reader may enjoy deriving the slightly different algorithms that come about when other sequencings are used.

We shall examine our sequencing in reverse order. It is easy to determine whether we need to push the center button. Since all the other buttons would have already

been pushed if needed, we push the center button iff we are not already in the desired winning pattern.

It is just as simple to determine whether we need to push a particular corner button. The only buttons that can toggle the state of a corner button are that corner button and the two adjacent edge buttons. In our ordering, if we want to push the edge buttons, we would have done so by the time that we consider the corner button. Therefore, we push a corner button iff it is not already in its proper state for the winning pattern.

It remains only to determine which edge buttons need to be pushed. We can use Pelletier's matrix to establish this. For example, consider the edge button numbered 2. This corresponds to the second row of Pelletier's matrix. Since ones occur in columns 1 through 6 of the matrix, Pelletier's algorithm may be interpreted as saying the following: we push button number 2 iff we want to change the parity of the number of flashing buttons among buttons 1 through 6.

We shall call buttons 1 through 6 the *dependency rectangle* of button number 2. In general, we define the dependency rectangle of an edge button to be the edge button and the five adjacent buttons; these six buttons form a 2 by 3 or 3 by 2 rectangular array. By symmetry, we see that an edge button should be pushed iff we want to change the parity of the number of flashing buttons in its dependency rectangle. Note that we may calculate in advance which edge buttons to push, or we may consider them in turn; pushing one edge button does not affect the parity of the other edge buttons' dependency rectangles. This is as it should be, since pushing any edge button does not affect whether we want to push any other edge button.

The parity arguments we have given work for any desired winning pattern. For the specified winning pattern, our algorithm becomes the following:

1) For each edge button in turn, push it iff there are an even number of flashing buttons in its dependency rectangle.

2) Push any corner buttons that are off.

3) If the winning pattern is not already displayed, push the center button.

REFERENCE

1. Donald H. Pelletier, Merlin's Magic Square, this MONTHLY, 94 (1987) 143–150.

## The Number of Words of Length $n$ in a Graph Monoid

DAVID C. FISHER*
*Department of Mathematics, Harvey Mudd College, Claremont, CA 91711*

**Introduction.** An old chestnut in combinatorics is:

**Old Problem.** How many $n$ letter words can be made from $m$ letters:
1. counting permutations (e.g., *caab* and *abac* are different)?
2. not counting permutations (e.g., *caab* and *abac* are the same)?

**Answers.** 1. $m^n$ 2. $\binom{n+m-1}{n}$

*Author's current address is Department of Mathematics, University of Colorado, Denver, CO 80204

The problem solved here includes these as special cases. Besides allowing no commuting or the commuting of all letters, we allow some pairs of letters to commute.

**New Problem.** How many $n$ letter words can be made from $m$ letters if certain pairs commute?

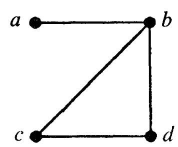
**Example.** How many four letter words can be made from $a$, $b$, $c$ and $d$ if $ab = ba$, $bc = cb$, $bd = db$ and $cd = dc$?

**Solution to example (hard way).** Do it by exhaustive search.

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *aaaa* | *aaab* | *aaac* | *aaad* | *aabb* | *aabc* | *aabd* | *aaca* | *aacc* | *aacd* | *aada* | *aadd* | *abbb* | *abbc* | *abbd* |
| *abca* | *abcc* | *abcd* | *abda* | *abdd* | *acaa* | *acac* | *acad* | *acca* | *accc* | *accd* | *acda* | *acdd* | *adaa* | *adac* |
| *adad* | *adda* | *addd* | *bbbb* | *bbbc* | *bbbd* | *bbca* | *bbcc* | *bbcd* | *bbda* | *bbdd* | *bcaa* | *bcac* | *bcad* | *bcca* |
| *bccc* | *bccd* | *bcda* | *bcdd* | *bdaa* | *bdac* | *bdad* | *bdda* | *bddd* | *caaa* | *caac* | *caad* | *caca* | *cacc* | *cacd* |
| *cada* | *cadd* | *ccaa* | *ccac* | *ccad* | *ccca* | *cccc* | *cccd* | *ccda* | *ccdd* | *cdaa* | *cdac* | *cdad* | *cdda* | *cddd* |
| *daaa* | *daac* | *daad* | *daca* | *dacc* | *dacd* | *dada* | *dadd* | *ddaa* | *ddac* | *ddad* | *ddda* | *dddd* | | |

So the answer is 88.

**Answer.** Form a graph with a node for each letter and edges between commuting letters. So the graph for the example is




Let $c_k$ be the number of $k$-complete subgraphs in the graph. (A $k$-complete subgraph is a set of $k$ nodes with an edge between each pair of nodes.) Let $T_n$ be the answer for a given $n$. We will show that for $n > 0$,

$$T_n = c_1 T_{n-1} - c_2 T_{n-2} + c_3 T_{n-3} - c_4 T_{n-4} + \cdots$$

with $T_0 = 1$ and for $n < 0$, $T_n = 0$.

**Solution to Example (Easy way).** There are four 1-complete subgraphs ($a$, $b$, $c$, $d$), four 2-complete subgraphs ($ab$, $bc$, $bd$, $cd$), one 3-complete subgraph ($bcd$) and no higher ones. Thus $T_n = 4T_{n-1} - 4T_{n-2} + T_{n-3}$ with $T_0 = 1$ and for $n < 0$, $T_n = 0$. So $T_1 = 4$, $T_2 = 12$, $T_3 = 33$ and $T_4 = 88$.

This note is devoted to verifying this result and solving the difference equation to obtain explicit formulas for $T_n$. While this result can be easily derived from the work of Cartier and Foata [1], this paper presents a more direct and less abstract approach.

**Properties of Graph Monoids.**

**Definitions and Trivial Results.** Let $G$ be a graph with nodes $V$ and edges $E$. Let $S$ be the free monoid generated by $V$. Let $\varnothing$ be the "null word". Let $p$, $q$, $r$, $s$, $t$,

$u, v, x, y \in S$. Then:

1. $x \sim y$ iff $x = ab$ and $y = ba$ with $(a, b) \in E$.
2. $x \approx y$ iff $x = upv$ and $y = uqv$ with $p \sim q$.
3. $x \doteq y$ iff $x = z_0 \approx z_1 \approx \cdots \approx z_k = y$ ($k$ may be zero, so $x \doteq x$). $\doteq$ is the congruence on $S$ generated by $\sim$.
4. The *graph monoid*, $R$ is the monoid generated from $V$ subject to $\doteq$.
5. $x \prec y$ iff $y \doteq rxs$. If $r = \varnothing$, then $x \ll y$. (If $x \prec y$ and $y \doteq z$, then $x \prec z$.)
6. $|x|$ is the *length* of $x$. (If $x \doteq y$, then $|x| = |y|$.)
7. $K_n(x) = \{w \in R$ such that $|w| = n$ and $x \ll w\}$ and $T_n = |K_n(\varnothing)|$.

The following is a string of technical lemmas necessary to prove Theorem 1. The reader may, without loss of continuity, skip to Theorem 1.

LEMMA 1. *Let $a \in V$. If $ras \doteq \tilde{r}a\tilde{s}$ with either $a \nprec r\tilde{r}$ or $a \nprec s\tilde{s}$, then $rs \doteq \tilde{r}\tilde{s}$.*

*Proof.* Assume $a \nprec r\tilde{r}$. If $a \nprec s\tilde{s}$, the proof is similar.

If $ras \sim \tilde{r}a\tilde{s}$, then either $r = \tilde{s} = \varnothing$ and $s = \tilde{r}$ with $(a, s) \in E$, or $s = \tilde{r} = \varnothing$ and $r = \tilde{s}$ with $(r, a) \in E$. Either way, $rs = \tilde{r}\tilde{s}$.

If $ras \approx \tilde{r}a\tilde{s}$, then $ras = upv$ and $\tilde{r}a\tilde{s} = uqv$. If $a \prec u$, then $u = wat$ with $a \nprec w$. Thus $\tilde{r} = w$ and $rs = wtpv \approx wtqv = \tilde{r}\tilde{s}$. If $a \nprec u$ but $a \prec p$, then since $p \sim q$, $a \prec q$. So $p = wat$ and $q = \tilde{w}a\tilde{t}$ with $wt \sim \tilde{w}\tilde{t}$. Thus $rs = uwtv \approx u\tilde{w}\tilde{t}v = \tilde{r}\tilde{s}$. Otherwise, $a \prec v$ and $a \nprec up$, then $v = wat$ with a $a \nprec upw \approx uqw$. So $a \nprec uqw$. Thus $\tilde{s} = t$ and $rs = upwt \approx uqwt = \tilde{r}\tilde{s}$.

If $ras \doteq \tilde{r}a\tilde{s}$, then $ras = t_0 \approx t_1 \approx \cdots \approx t_k = \tilde{r}a\tilde{s}$. Since $a \prec x$ and $x \doteq t_j$, $a \prec t_j$. So $t_j = r_j a s_j$ with $a \nprec r_j$. So $rs = r_0 s_0 \approx r_1 s_1 \approx \cdots \approx r_k s_k = \tilde{r}\tilde{s}$.

LEMMA 2. *If $xy \doteq x\tilde{y}$, then $y \doteq \tilde{y}$. Also, if $xy \doteq \tilde{x}y$, then $x \doteq \tilde{x}$.*

*Proof.* Only the first statement is proved. The proof the second statement is similar.

This is done by induction on $|x|$. For $|x| = 0$, it is trivial. For $|x| > 0$, factor $x = as$ with $a \in V$. By Lemma 1 with $r = \varnothing$, $sy \doteq s\tilde{y}$. Since $|s| = |x| - 1$, induction gives $y \doteq \tilde{y}$.

LEMMA 3. *If $a \in V$, $a \ll x$, $t \ll x$ and $a \not\ll t$, then $at \ll x$ and $at \doteq ta$.*

*Proof.* Let $au \doteq x \doteq tv$. First, we need to show that $a \nprec t$. Suppose not. Then $t = ras$ with $a \nprec r$. Then $au \doteq rasv$, so Lemma 1 implies $u \doteq rsv$. Thus, $rasv \doteq au \doteq arsv$. By Lemma 2, $t = ras \doteq ars$. So $a \ll t$, which is a contradiction.

Since $a \prec tv$ and $a \nprec t$, $a \prec v$. Factor $v \doteq ras$ with $a \nprec r$. By Lemma 1, $u \doteq trs$. So $au \doteq atrs$. Thus $at \ll x$ and $atrs \doteq tras \Rightarrow atr \doteq tra$.

We show $at \doteq ta$ by induction on $|r|$. For $|r| = 0$, it is trivial. For $|r| > 0$, factor $r = wb$ with $b \in V$. Since $a \nprec r$, $a \neq b$. By Lemma 1, $atw \doteq twa$. Since $|w| = |r| - 1$, induction gives $at \doteq ta$.

LEMMA 4. *Let $A \equiv \{a_1, a_2, \ldots, a_k\} \subseteq V$ and $w_A = a_1 a_2 \cdots a_k$. If for all $a \in A$, $a \ll x$, then $w_A \ll x$ and for all $a, b \in A$, $ab \sim ba$ i.e. $A$ is a complete subgraph.*

*Proof.* Suppose $c, d \in A$ with $cd \nsim dc$. But since $c \ll x$, $d \ll x$ and $c \not\ll d$ this violates Lemma 3. So $A$ is a complete subgraph.

We show $w_A \ll x$ by induction on $k$. If $k = 0$, it is trivial. For $k > 0$, let $B \equiv A - \{a_k\}$. Then by induction $w_B \ll x$. Since $a_k \ll x$ and $a_k \not\ll w_B$, $a_k w_B \doteq w_A \ll x$.

LEMMA 5. *For* $A \neq \varnothing$,

$$\bigcap_{a \in A} K_n(a) = \begin{cases} K_n(w_A) & \text{if } A \text{ is a complete subgraph} \\ \varnothing & \text{otherwise} \end{cases}$$

*Proof.* Since $K_n(w_A) \subseteq K_n(a)$ for all $a \in A$, $K_n(w_A) \subseteq \bigcap_{a \in A} K_n(a)$. Let $x \in \bigcap_{a \in A} K_n(a)$. Then by Lemma 4, $A$ is a complete subgraph and $w_A \ll x$. So $x \in K_n(w_A)$ and hence $\bigcap_{a \in A} K_n(a) \subseteq K_n(w_A)$. If $A$ is not a complete subgraph, then $\bigcap_{a \in A} K_n(a)$ must be empty.

LEMMA 6. $|K_n(p)| = |K_{n-|p|}(\varnothing)|$.

*Proof.* The mapping $F\colon K_{n-|p|}(\varnothing) \Rightarrow K_n(p)$ with $f_G(x) \equiv px$ is onto. By Lemma 2, it is also one-to-one.

**Formulas for $T_n$.** Three methods for finding the number of words of length $n$ in a graph monoid are given. Theorem 1 gives a difference equation for $T_n$. This difference equation is solved in Corollary 2 to give two formulas for $T_n$. One involves taking the $n$th derivative of a rational function, while the other is a closed form solution in terms of the roots of a polynomial.

THEOREM 1. *For all* $n > 0$, $T_n = \sum_{k=1}^{n}(-1)^{k+1}c_k T_{n-k}$ *with* $T_0 = 1$.

*Proof.* Let $J$ be the set of nonempty complete subgraphs in $G$. Since $n > 0$ and using the inclusion-exclusion principle and Lemmas 5 and 6:

$$T_n = |K_n(\varnothing)| = \left|\bigcup_{a \in V} K_n(a)\right| = \sum_{\varnothing \neq A \subseteq V} (-1)^{|A|+1}\left|\sum_{a \in A} K_n(a)\right|$$

$$= \sum_{A \in J} (-1)^{|A|+1}|K_n(w_A)| = \sum_{A \in J} (-1)^{|A|+1}|K_{n-|A|}(\varnothing)| = \sum_{k=1}^{n} (-1)^{k+1} c_k T_{n-k}.$$

DEFINITION. *Let $G$ be a graph. Then the clique polynomial of $G$ is* $f_G(z) \equiv 1 - c_1 z + c_2 z^2 - c_3 z^3 + \cdots$ *where $c_k$ is the number of $k$-complete subgraphs of $G$.*

COROLLARY 1. $f_G(z)(T_0 + T_1 z + T_2 z^2 + \cdots) = 1$.

*Proof.* Let $a_0 + a_1 z + a_2 z^2 + \cdots \equiv f_G(z)(T_0 + T_1 z + T_2 z^2 + \cdots)$. Clearly, $a_0 = T_0 = 1$. For $n > 0$, Theorem 1 gives $a_n = T_n - \sum_{k=1}^{n}(-1)^{k+1}c_k T_{n-k} = 0$.

COROLLARY 2. *If $z_1, z_2, \ldots, z_h$ are the roots of $f_G(z) = 0$ with multiplicities $l_1, l_2, \ldots, l_h$, then for $n \geqslant 0$:*

$$T_n = \frac{d^n}{dz^n}\left(\frac{1}{n! f_G(z)}\right)_{z=0} = -\sum_{j=1}^{h} \lim_{z \to z_j} \frac{d^{l_j - 1}}{dz^{l_j - 1}} \frac{(z - z_j)^{l_j}}{(l_j - 1)! z^{n+1} f_G(z)}.$$

*Proof.* The first expression follows from Corollary 1. So all that needs to be shown is that the second expression equals the first.

For all $n \geqslant 0$, $g_n(z) \equiv 1/z^{n+1}f_G(z)$ is a rational function. Since $\deg(z^{n+1}f_G(z)) \geqslant 2$, the residue at $\infty$ of $g_n(z)$ is zero. So by Theorem 2.6 of Markushevich

[4, vol. 2],

$$0 = \operatorname*{Res}_{x=0}(g_n(z)) + \sum_{j=1}^{h} \operatorname*{Res}_{z=z_j}(g_n(z))$$

$$= \frac{d^n}{dz^n}\left(\frac{1}{n!f_G(z)}\right)_{z=0} + \sum_{j=1}^{h} \lim_{z\to z_j} \frac{d^{l_j-1}}{dz^{l_j-1}} \frac{(z-z_j)^{l_j}}{(l_j-1)!z^{n+1}f_G(z)}.$$

REFERENCES

Corollary 1 easily follows from a theorem of Cartier and Foata [1, pp. 1, 8–23]. However, they did not use their result to count anything. Nor did they use graphs in the way they are used here. Because Cartier and Foata's hypothesis is less restrictive and their result more general, their proof is more difficult than the proof given here.

Also, Kim *et al.* [2] derived properties of the graph monoid. Their graph is the complement of the one used here.

1. P. Cartier and D. Foata, Problèmes Combinatoires de Commutation et Rèarrangements, Lecture Notes in Mathematics, 85, Springer-Verlag, Berlin, 1969.
2. K. H. Kim, L. G. Makar-Limanov, and F. W. Roush, *Graph Monoids*, Semigroup Forum, 25(1982), 1–7.
3. L. Lovâsz, Combinatorical Problems and Exercises, North-Holland Publishing Company, New York, 1979.
4. A. I. Markushevich, Theory of Functions of a Complex Variable, Chelsea Publishing Company, New York, 1977.

# On the Number of Furthest Neighbour Pairs in a Point Set

HERBERT EDELSBRUNNER[1]
*Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801*

STEVEN S. SKIENA
*Department of Computer Science, State University of New York, Stony Brook, NY 11794*

**Introduction.** Consider a configuration $N$ of points in the Euclidean plane, labeled $1, 2, \dots, n$. A point $j$ is a *furthest neighbour* of $i$ if $d(i, j) = \max_{1 \leqslant k \leqslant n}\{d(i, k)\}$, where $d$ is the Euclidean distance function. An ordered pair of points $(i, j)$, $i, j \in N$ is a *furthest neighbour pair* of $N$ if $j$ is a furthest neighbour of $i$. We let $M(N)$ denote the number of furthest neighbour pairs defined by $N$. Each point has at least one furthest neighbour, so we have $M(N) \geqslant n$. Our interest is in $M(n)$, the largest number of furthest neighbour pairs possible in a configuration of $n$ points. Avis [1] proves $M(n) = 3n - 3$ for even $n \geqslant 4$ and $3n - 4 \leqslant M(n) \leqslant 3n - 3$ for odd $n \geqslant 5$. Counting furthest neighbour pairs can be generalized to counting repeated distances with other restrictions, as discussed in [2].

This note corrects Avis' proof for the even case and improves the result to $M(n) = 3n - 4$ for odd $n \geqslant 5$. Also, we show that a convex set of $n \geqslant 3$ points has at most $2n$ furthest neighbour pairs and that this bound is tight. Finally, we prove

[1]Research supported by Amoco Fnd. Fac. Dev. Comput. Sci. 1-6-44862.

that the number of furthest neighbour pairs of $n$ points in three dimensions is subquadratic if no three points are collinear; this is the case if the set is convex.

**Results in Two Dimensions.** First, we introduce some terminology which borrows from [**1**]. Let $n_i$ be the number of furthest neighbours of $i$ and let $m_i$ be the number of points $j$ such that $(j, i)$ is a furthest neighbour pair. We use the distance $r_i$ of $i$ to its furthest neighbours as a radius to define circle $C_i$ and disk $D_i$ centered at $i$. Each disk $D_i$ contains all points $j \in N$, so $N$ lies in $\bigcap_{i=1}^{n} D_i$. Thus $n_i$ can also be defined as the number of points on $C_i$ and $m_i$ as the number of circles passing through $i$.

Avis proves $M(N) \leqslant 3n - 3$ by considering the circles $C_i$, $i = 1, 2, \ldots, n$, in order of non-decreasing radius, that is, $r_i \leqslant r_{i+1}$. Each point that lies on $C_k$ is one of two types, either one which lies on a $C_j$, $j < k$ or one which does not. We let $h_k$ be the number of points of the first type and $f_k$ the number of the second type. By definition,

$$M(N) = \sum_{k=1}^{n} f_k + \sum_{k=2}^{n} h_k.$$

Since each point can be intersected for the first time only once, $\sum_{k=1}^{n} f_k \leqslant n$. This coupled with Lemma 1 gives $M(N) \leqslant 3n - 2$, which can be improved to the stated result.

Unfortunately, the original proof of Lemma 1 given in [**1**] uses an incorrect argument. We present an alternate proof:

LEMMA 1. $rh_k \leqslant 2$ *for* $k = 2, 3, \ldots, n$.

*Proof.* Assume $h_k \geqslant 3$. A point is of the first type with respect to $C_k$ only if it lies on the boundary $B$ of $\bigcap_{i=1}^{k-1} D_i$. Thus $C_k$ intersects $B$ in at least three points. Since $r_i \leqslant r_k$ for all $i < k$, at each intersection at least one side of $C_k$ is outside $B$. Consequently, the intersection points divide $C_k$ into arcs at least two of which lie outside $B$. The smallest of these arcs spans an angle $\theta < \pi$ with respect to $k$. Thus, there is an index $i < k$ such that $C_i$ intersects $C_k$ in two points that belong to this arc. So an arc of $C_k$ spanning an angle greater than $\pi$ must be within $C_i$. This is a contradiction since $r_i \leqslant r_k$.

We use two results to prove our main theorem. The first is a technical lemma, the other a bound on $M(N)$ for convex sets $N$. In this context, a set of points is said to be *convex* if each point is a vertex of the convex hull of the set.

LEMMA 2. *For a set* $N$ *of* $2k$ *points on a circle* $C$, *we have* $M(N) < 4k$.

*Proof.* Assume the converse, that there exists a circular arrangement $N$ of $2k$ points where $M(N) \geqslant 4k$. Since any two different circles intersect in at most two points, each point has at most two furthest neighbours on $C$. Thus $M(N) = 4k$.

Let $u$ and $v$ be furthest neighbours of $s$ on $C$ and $c$ be the center of $C$. When the points are angularly ordered around $c$, $u$ and $v$ must be consecutive points. If there were to be another point $w$ between them, either $d(s, w) > d(s, v)$ or $d(s, u) \neq d(s, v)$. If $M(N) = 4k$, a point $u$ can be furthest neighbour of another point only if this other point has another furthest neighbour $v$. Now, $v$ can only be the predecessor or the successor of $u$. Consequently, $u$ is furthest neighbour of exactly two points. These two points, $s$ and $t$, are also the furthest neighbours of $u$.

Otherwise, $u$ has a furthest neighbour $r$ between $s$ and $t$ which implies that $u$ is also furthest neighbour of $r$, a contradiction. As a consequence, the points must be equally spaced around $C$. However, the line through $s$ and $c$ will intersect another point if an even number of points are equally spaced, so $s$ has only one furthest neighbour, a contradiction. Therefore, $M(N) < 4k$.

THEOREM 3. *For a convex set $N$ of $n \geqslant 3$ points, $M(N) \leqslant 2n$ and the bound is tight.*

*Proof.* Label the points according to their counterclockwise order around the convex hull of $N$. If $v_1, v_2, \dots, v_m$ are the successive furthest neighbours of point $i$, we show that only $v_m$ can be a furthest neighbour of $i + 1$. Consider the perpendicular bisector between $v_1$ and $v_2$, which passes through $i$. Point $i + 1$ is on the same side of the bisector as $v_1$, which means $d(i + 1, v_1) < d(i + 1, v_2)$ and similarly $d(i + 1, v_j) < d(i + 1, v_{j+1})$, $j < m$.

We maintain two pointers as we move around $N$, $i$ and $v$, where $v$ is a furthest neighbour of $i$. Initially $i = 1$ and $v$ is the first furthest neighbour of $i$. Each step moves $v$ forward until it is the $n_i{}^{\text{th}}$ furthest neighbour, when $i$ is advanced. Neither $i$ nor $v$ can retreat. We stop when $i = n$ and $v$ is the $n_n{}^{\text{th}}$ neighbour of $n$, which must be less than or equal to the initial value of $v$. The situation is shown in Figure 1. Each step corresponds to a furthest neighbour pair, and the only pair not defined by a step is represented by the initial configuration. Since $i$ is advanced $n - 1$ times, and $v$ at most $n$ times, adding the initial pair gives $M(N) \leqslant n + (n - 1) + 1 = 2n$.

To show that the upper bound is tight, take the vertices of an equilateral triangle and, for each vertex $i$, draw the shorter circular arc centered at $i$ that connects the other two vertices. Pick the three vertices together with $n - 3$ arbitrary points on the three arcs to form $N$; then $M(N) = 2n$.

With these results we can prove the exact value of $M(n)$. For even $n$ the result is the same as in Avis [**1**]. For completeness and since it is requires no extra effort we include this case.

THEOREM 4. *$M(n) = 3n - 3$ if $n \geqslant 4$ is even and $M(n) = 3n - 4$ if $n \geqslant 5$ is odd.*

*Proof.* We perform a case analysis on $m$, the number of points that are not vertices of the convex hull of $N$.

(1) Let $m = 0$. The $n$ points form a convex set. By Theorem 3, $M(N) \leqslant 2n$.

(2) Let $m = 1$. Exactly one point $i$ violates convexity. In this case, $i$ cannot be furthest neighbour of any other point. Either $n_i = n - 1$ or $n_i < n - 1$. The first possibility implies the $n - 1$ points lie on a circle. Thus $M(N) \leqslant n - 1 + 2(n - 1)$ which implies the result for even $n$. If $n$ is odd, then by Lemma 2, $M(N) < n_i + 2(n - 1) \leqslant 3n - 3$. Similarly, if $n_i < n - 1$ and the rest of the points are convex, by Theorem 3:

$$M(N) \leqslant n_i + 2(n - 1) \leqslant n - 2 + 2(n - 1) \leqslant 3n - 4.$$

(3) Let $m \geqslant 2$. A total of $m$ points violate convexity. Since none of the $m$ points can be the furthest neighbour of any other point,

$$M(N) = \sum_{k=1}^{n} n_k = \sum_{k=1}^{n} f_k + \sum_{k=2}^{n} h_k \leqslant (n - m) + 2(n - m) \leqslant 3n - 4.$$

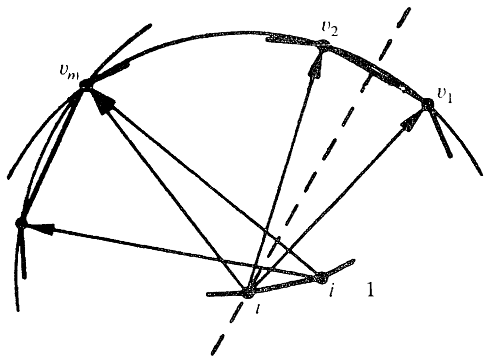



FIG. 1. Counting furthest neighbours in a convex set.

A point set $N$ realizing this bound can be constructed. If $n$ is even arrange $n - 1$ points equally spaced along a circle, one point $c$ in the center. The center point has $n - 1$ furthest neighbours, and the $n - 1$ points have two furthest neighbours each. Thus, $M(N) = 3n - 3$. If $n$ is odd add a point along a radius between $c$ and any other point to the construction for $n - 1$ points. Now, the center point has $n - 2$ furthest neighbours, and again every other point has two furthest neighbours; so $M(N) = 3n - 4$.

**Extensions to Three Dimensions.** In three dimensional Euclidean space, we can construct a set $N$ of $n$ points with $M(N) \geqslant (n^2 + 2n)/4$. Choose half of the points on a circle $C$ and choose the other points on the perpendicular line $l$ through the center of $C$. If $n$ is odd, then let the number of points on $C$ be one more than on $l$. The points on $l$ can be picked such that for each $i$ on $l$ and for each $j$ on $C(i, j)$ is a furthest neighbour pair. Only the lower order terms of this construction can be improved (see [2]).

Note that the above construction uses about $n/2$ points which are collinear. Interestingly, collinear points are necessary to obtain a quadratic number of furthest neighbour pairs.

THEOREM 5. *$M(N) \leqslant c \cdot n^{5/3}$ if $N$ is a set of $n$ points in three dimensions such that no three points are collinear.*

*Proof.* Let $r_i$ be the distance from point $i$ to its furthest neighbours and let $S_i$ be the sphere with radius $r_i$ and center $i$. Three such spheres intersect in at most two points since their centers are not collinear.

Define the 3-regular multi-hypergraph $H$ with node set $N$ that contains a hyperedge $\{i, j, k\}$ $m$ times if points $i$, $j$, and $k$ share $m$ common furthest neighbours. Since $|S_i \cap S_j \cap S_k| \leqslant 2$, each hyperedge can occur at most twice which implies that $H$ has at most $2\binom{n}{3}$ hyperedges. Recall that point $i$ is furthest neighbour of $m_i$ other points. Thus, $i$ contributes $\binom{m_i}{3}$ hyperedges to $H$. This implies

$$\sum_{i=1}^{n} \binom{m_i}{3} \leqslant 2\binom{n}{3}.$$

By the Cauchy-Schwartz inequality (see [3]), we infer

$$\sum_{i=1}^{n} m_i \leqslant 2^{1/3} n^{5/3} + o(n^{5/3}).$$

For example, if $N$ is a convex set then no three of its points are collinear and therefore, $M(N) = O(n^{5/3})$. It is not known whether or not $M(N)$ can be superlinear in this case.

**Acknowledgement.** We would like to thank an anonymous referee for suggestions that improved the readability of this paper.

REFERENCES

1. D. Avis, The number of furthest neighbour pairs of a finite planar set, *Amer. Math. Monthly*, 91 (1984) 417–420.
2. D. Avis, P. Erdös, and J. Pach, Repeated distance in space, manuscript, School Comput. Sci., McGill Univ., 1986.
3. P. R. Halmos, Finite-dimensional Vector Spaces, 2nd edition, D. Van Nostrand, Princeton, NJ, 1958.

# A Remark on Divided Differences

E. T. Y. LEE
*Boeing Commercial Airplane Co., P.O. Box 3707, M/S 6E-27, Seattle, WA 98124*

**1. Introduction.** In a recent article [6] in this MONTHLY, Saff and Snader presented a new proof of a formula due to Schneider [7], expressing the error for a general quadrature formula. Their proof is based on the contour integral representation of divided differences and the use of complex variable techniques, which, in their words, "simplify the algebraic manipulation of divided differences, and provide a straightforward derivation" of the error formula. While the complex variable approach is illuminating, at least pedagogically, one purpose of this note is to point out that the same result can be obtained directly, without complex variables, in an even more straightforward manner. The direct proof depends on a simple identity, which, though not new, does not seem to belong to the familiar arsenal for the manipulation of divided differences. Thus, a second purpose of the note is to draw attention to this identity, by presenting several other examples of its applications.

Schneider's formula will be given in Section 3. The proof we constructed on reading [6] turns out to be essentially identical to Schneider's original. However, since his thesis is not readily accessible to the general readership, a repetition of this short proof may not be out of place. I wish to thank Claus Schneider for sending me a copy of his thesis.

**2. The divided difference.** Divided differences are basic in certain parts of numerical analysis. For instance, they are intimately connected with Newton's interpolation formula. More recently, they are also directly involved in the definition of $B$-splines. Among the several equivalent definitions, I prefer the following more conceptual one:

DEFINITION. *The nth order divided difference, $f[x_0,\dots,x_n]$, of a function $f(x)$ at the (real or complex) points $x_0,\dots,x_n$, is the coefficient of $x^n$ of the polynomial $p(x)$, degree at most $n$, which agrees with (i.e., interpolates to) $f(x)$ at $x = x_0,\dots,x_n$.*

This definition, rather than the more traditional one using the recursion (2) below, or the more concrete one in terms of ratios of determinants, generally allows for more efficient and elegant developments of their properties; see for instance [**1**], [**2**]. The arguments in the definition need not be distinct: Two functions are said to agree at a point $j$ times if and only if their values and first $j-1$ derivatives are equal there. In such a situation we will assume sufficient smoothness of the functions at the relevant points. Behind this definition, of course, is the theorem that there is a unique polynomial, degree at most $n$, that satisfies these *osculatory* interpolation conditions, a proof of which can be found in [**3**, p. 29]. This interpretation will be used throughout. Thus, for instance, $f[x_0, x_0] = f'(x_0)$.

It is immediate from the definition that $f[x_0,\dots,x_n]$ is symmetric in its arguments, and that if $f$ and $g$ agree at $x_0,\dots,x_n$, then

$$f[x_0,\dots,x_n] = g[x_0,\dots,x_n].$$

The *remark* of the title is the following trivial observation: If $p$ is a polynomial agreeing with $f$ at $x_0,\dots,x_n$, and if $f$ is also defined at $x_{n+1}$, then $(x - x_{n+1})p(x)$ agrees with $(x - x_{n+1})f(x)$ at $x = x_0,\dots,x_{n+1}$. (If $x_{n+1}$ already appears $j$ times in the list $x_0,\dots,x_n$, the new functions will agree $j+1$ times there, as is immediate from Leibniz' formula for differentiation.) Hence

$$f[x_0,\dots,x_n] = ((\cdot - x_{n+1})f)[x_0,\dots,x_{n+1}]. \tag{1}$$

The usefulness of this identity is that it allows one to expand or shrink the argument list in a divided difference. To illustrate, we note that the basic recurrence formula for divided differences,

$$(x_n - x_0)f[x_0,\dots,x_n] = f[x_1,\dots,x_n] - f[x_0,\dots,x_{n-1}] \tag{2}$$

(which, as it stands, is useful only when $x_n \neq x_0$), results immediately upon subtraction of the following two applications of (1):

$$f[x_1,\dots,x_n] = ((\cdot - x_0)f)[x_0,\dots,x_n],$$
$$f[x_0,\dots,x_{n-1}] = ((\cdot - x_n)f)[x_0,\dots,x_n].$$

It should be noted that the identity (1) is simply a special case of the Leibniz formula for divided differences, (see (7) below). On the other hand, this special case also provides a short proof of the full formula, as we will show in Section 4. This proof seems new.

While preparing this article, we found that (1) also appears in [**4**], together with the same approach to (2). The argument leading to (1) is somewhat different there: the points $x_0,\dots,x_n$ are distinct, and (1) follows from an examination of the Lagrange interpolation formula.

**3. Schneider's Error Formula.** Given a quadrature formula

$$Q_n(f) := \sum_{j=0}^{n} \alpha_j f(x_j) \tag{3}$$

for $\int_a^b f(x)\,dx$, (where $x_0, \ldots, x_n$ are distinct reals that include $a$, $b$), the error

$$E_n(f) := \int_a^b f(x)\,dx - Q_n(f) \tag{4}$$

can be expressed as a divided difference of order $2n+1$, each $x_i$ occurring twice in the arguments:

THEOREM (SCHNEIDER). *Let $F$ be any primitive of $f$, that is, $F' = f$, and let*

$$\Omega(x) := \prod_{j=0}^{n} (x - x_j),$$

$$q(x) := \Omega^2(x)\left(\frac{1}{x-b} - \frac{1}{x-a} - \sum_{j=0}^{n} \frac{\alpha_j}{(x-x_j)^2}\right). \tag{5}$$

*Then the error* (4) *of the formula* (3) *is*

$$E_n(f) = (Fq)[x_0, x_0, \ldots, x_n, x_n]. \tag{6}$$

*Proof.* $E_n(f) = F(b) - F(a) - \Sigma_0^n \alpha_j F'(x_j)$. Repeated application of (1) gives

$$F(a) = F[a] = ((\cdot - a)F)[a, a] = \left(\frac{\Omega^2 F}{(\cdot - a)}\right)[x_0, x_0, \ldots, x_n, x_n]$$

and

$$F'(x_j) = F[x_j, x_j] = \left(\frac{\Omega^2 F}{(\cdot - x_j)^2}\right)[x_0, x_0, \ldots, x_n, x_n],$$

from which (5), (6) follow immediately.

As mentioned, this proof is identical to Schneider's. However, in [7], the identity (1), or rather a slightly general form of it, (with several added arguments instead of one), is proved using the Leibniz formula plus the general form of the divided difference (in terms of values and derivatives). This is perhaps what leads Saff and Snader to say that Schneider's proof, "which makes use of Leibniz's rule for divided differences and other identities, is algebraic in nature", and to assert in effect that theirs is simpler.

**4. Leibniz' Formula.** This is a formula for the divided difference of a product of two functions, which reduces to the better known formula for differentiation in case all arguments coalesce. (Hence Leibniz' name.) It says

$$(fg)[x_0, \ldots, x_n] = \sum_{k=0}^{n} f[x_0, \ldots, x_k]\, g[x_k, \ldots, x_n]. \tag{7}$$

The usual proof of (7) by induction is a bit lengthy and not too exciting. In [**1**, p. 5] there is an interesting argument attributed to W. D. Kammler which is, however, somewhat involved and, at least on first reading, not particularly easy to understand. We feel that the following short proof based on (1) is much more transparent.

*Proof.* Let $p$ be the polynomial of degree at most $n$ agreeing with $f$ at $x_0, \dots, x_n$. In its Newton form, (see for instance [**1**, p. 4]),

$$p(x) = \sum_{k=0}^{n} f[x_0, \dots, x_k](x - x_0) \cdots (x - x_{k-1}).$$

Since $pg$ agrees with $fg$ at $x_0, \dots, x_n$, we have

$$\begin{aligned}(fg)[x_0, \dots, x_n] &= (pg)[x_0, \dots, x_n] \\ &= \sum_{k=0}^{n} f[x_0, \dots, x_k]((\cdot - x_0) \cdots (\cdot - x_{k-1})g)[x_0, \dots, x_n] \\ &= \sum_{k=0}^{n} f[x_0, \dots, x_k] g[x_k, \dots, x_n],\end{aligned}$$

the last line from repeated applications of (1).

One might ask, what is the polynomial $P$ of degree at most $n$, agreeing with the product $fg$ at $x_0, \dots, x_n$? Let $G_{n-k}$ be the polynomial, degree at most $n - k$, agreeing with $g$ at $x_k, \dots, x_n$. A slightly subtler reasoning will show that

$$P(x) = \sum_{k=0}^{n} f[x_0, \dots, x_k](x - x_0) \cdots (x - x_{k-1}) G_{n-k}(x). \tag{8}$$

Thus we have a second proof of Leibniz' formula: Simply read off the leading coefficients in (8). Once stated, (8) is easy to verify, so we shall leave it as an "exercise for the reader". (I remind the reader again that the points $x_0, \dots, x_n$ are not necessarily distinct.)

**5. Another Example.** The following identity is proved by Micchelli [**5**], as a preliminary step in obtaining a recurrence formula for multivariate $B$-splines. It says

$$\begin{gathered} f[x_0, \dots, x_n] = \sum_{k=0}^{n} \lambda_k f[\xi, x_0, \dots, x_{k-1}, x_{k+1}, \dots, x_n], \\ \xi = \sum_0^n \lambda_k x_k, \ \sum_0^n \lambda_k = 1. \end{gathered} \tag{9}$$

We note that (9) also follows easily from (1):

$$\begin{aligned} f[\xi, x_0, \dots, x_{k-1}, x_{k+1}, \dots, x_n] &= ((\cdot - x_k) f)[\xi, x_0, \dots, x_n], \\ \sum_0^n \lambda_k f[\xi, x_0, \dots, x_{k-1}, x_{k+1}, \dots, x_n] &= \left( \sum_0^n \lambda_k (\cdot - x_k) f \right)[\xi, x_0, \dots, x_n] \\ &= ((\cdot - \xi) f)[\xi, x_0, \dots, x_n] \\ &= f[x_0, \dots, x_n]. \end{aligned}$$

REFERENCES

1. C. de Boor, A Practical Guide to Splines, Springer-Verlag, New York, 1978.
2. S. D. Conte and C. de Boor, Elementary Numerical Analysis, third edition, McGraw-Hill, New York, 1980.

# On $(\mathbf{a} \times \mathbf{b}) \times \mathbf{c}$

RONALD SHAW AND FRED J. YEADON
*School of Mathematics, University of Hull, Hull, HU6 7RX, England*

**1. Introduction.** Of the basic results in elementary 3-dimensional vector algebra, probably the most intricate is the identity

$$(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{b} \cdot \mathbf{c})\mathbf{a} \tag{1}$$

for the repeated vector cross product. This vector-valued identity is easily seen to be completely equivalent to the scalar-valued identity

$$(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) = (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}) - (\mathbf{a} \cdot \mathbf{d})(\mathbf{b} \cdot \mathbf{c}). \tag{2}$$

In teaching vector algebra to students the problem arises of the "best" method of proving (1) or, equivalently, (2). In this note we list several different proofs, some ancient and some modern. (Just possibly the proofs D and F may be new.) After each proof we comment on its virtues and defects—mostly from the teaching point of view, but in some cases we discuss the links with more advanced mathematics.

Before embarking upon these proofs of the identities (1) and (2), we have to address the definition of $\mathbf{a} \times \mathbf{b}$. On the whole, from the teaching point of view, it seems best to adopt the traditional approach, in which a nonrigorous knowledge of geometry is assumed and $\mathbf{a} \times \mathbf{b}$ is defined via $\|\mathbf{a}\|\,\|\mathbf{b}\| \sin\theta\ \mathbf{n}$, with the sign of the unit vector $\mathbf{n}$ being fixed by a "right-hand rule." However, two algebraic definitions of $\mathbf{a} \times \mathbf{b}$ deserve mention. Let $E$ denote a real, 3-dimensional, oriented vector space, equipped with inner product $\mathbf{a} \cdot \mathbf{b}$. For the first algebraic definition (see [**4**, p. 129]) let $\Delta$ denote that determinant function (= trilinear alternating form) for $E$ which takes the value $+1$ upon some (indeed, any) positive orthonormal basis: $\Delta(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = +1$. Then $\mathbf{a} \times \mathbf{b}$ is defined to be that vector such that

$$(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = \Delta(\mathbf{a}, \mathbf{b}, \mathbf{c}) \tag{3}$$

for all $\mathbf{c} \in E$. The properties of $\mathbf{a} \times \mathbf{b}$ then follow from those of $\Delta$ and of the scalar product. In particular the crucial bilinearity property of $\mathbf{a} \times \mathbf{b}$ (which is tricky to prove in some approaches) follows from the trilinearity of $\Delta$ and the bilinearity of $\mathbf{a} \cdot \mathbf{b}$.

*Remark.* Traditionally the logic proceeds in the opposite direction, (3) being used as the *definition* of the *scalar triple product* $\Delta$ in terms of previously defined dot and cross products. As one of us has argued elsewhere [**8**], this traditional order is surely misconceived, in that it induces the false belief that $\Delta$ is highly metrical in nature, being dependent upon two metrical constructs, namely the dot and cross products. In fact, of course, $\Delta$ does not depend upon notions of length and angle—as witness its definition (up to normalization convention) as a determinant function. (Certainly for $\dim E > 3$ no one seriously considers introducing determinant functions by a generalization of the metrical equation (3)!) Nevertheless, for the student's first encounter with vector algebra it seems best to keep clear of algebraic approaches,

and instead to give $\Delta$ a geometric definition in terms of oriented volume. This should be done *before* the introduction of the vector cross product, and the nonmetrical nature of volume can be stressed. The properties of $\Delta$, including trilinearity, are very simply obtained from this geometric definition (see [**8**]). After $\mathbf{a} \times \mathbf{b}$ has been defined in terms of vector area, as previously, then (3) can be derived via the metrical result (area of base) $\times$ height $=$ volume, and the bilinearity of $\mathbf{a} \times \mathbf{b}$ deduced.

The second algebraic approach is the axiomatic one, in which the real inner product space $E$ is equipped with a mapping $E \times E \to E : (\mathbf{a}, \mathbf{b}) \mapsto \mathbf{a} \times \mathbf{b}$ which satisfies the axioms

(X1) $\mathbf{a} \times \mathbf{b}$ is a bilinear function of $\mathbf{a}$ and $\mathbf{b}$.
(X2) $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{a} = 0 = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{b}$.
(X3) $(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{a} \times \mathbf{b}) = (\mathbf{a} \cdot \mathbf{a})(\mathbf{b} \cdot \mathbf{b}) - (\mathbf{a} \cdot \mathbf{b})^2$.

It is not hard to check that axioms (X1)–(X3) have, for $\dim E = 3$, only two solutions ("right handed" and "left handed")—those obtained via the two possible choices $+\Delta$ and $-\Delta$ of a normalized determinant function in (3). We will come back to this axiom system when discussing Proof F below.

**2. Proofs of the identities (1) and (2).**

*Proof A.* Step 1: From the geometric definition of the cross product, the vector $(\mathbf{a} \times \mathbf{b}) \times \mathbf{c}$ is perpendicular to $\mathbf{c}$ and (ignoring the trivial case when $\mathbf{a}$ and $\mathbf{b}$ are linearly dependent) lies in the plane of $\mathbf{a}$ and $\mathbf{b}$; consequently

$$(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} = \lambda(\mathbf{a}, \mathbf{b}, \mathbf{c})\{(\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{b} \cdot \mathbf{c})\mathbf{a}\} \tag{4}$$

for some scalar function $\lambda$ (defined for those $\mathbf{a}, \mathbf{b}, \mathbf{c}$ such that $(\mathbf{a} \cdot \mathbf{c})\mathbf{b} \neq (\mathbf{b} \cdot \mathbf{c})\mathbf{a}$).

Step 2: Show that $\lambda$ can be taken to be a constant function—whence (1) follows upon checking, for example, that $\lambda(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_1) = 1$.

*Comments.* Over the decades innumerable proofs have been given along these lines (see, e.g., [**1**, **3**]). The geometric part of the proof (Step 1) is both pleasing and promising, and, at first sight, Step 2 should be easy to accomplish. For, one reasons, surely $\lambda$ in (4) must be a constant function—since otherwise would not a conflict arise with the undoubted trilinearity of $(\mathbf{a} \times \mathbf{b}) \times \mathbf{c}$? Nevertheless Step 2 does seem to be irritatingly difficult to clinch, and one's pleasure in Proof A accordingly begins to evaporate. There is a danger, not always avoided, of trying to prove too much—for if in (4) one considers $\lambda$ as defined for all $\mathbf{a}, \mathbf{b}, \mathbf{c}$ then $\lambda$ is *not* forced to be a constant function. For example we could take $\lambda(\mathbf{a}, \mathbf{a}, \mathbf{c})$ to be any function of $\mathbf{a}$ and $\mathbf{c}$.

Many ways of passing from (4) to (1) (see, e.g., [**1**, **3**]) appeal to special features of the situation in hand, for example, symmetry properties. However, these special features are surely irrelevant and so one cries out for a general theorem to which one can appeal. Such a theorem is proved in the Appendix. However, even though the proof of this theorem is quite elementary, it must be admitted that students may well view it as an alien intrusion into their studies of 3-dimensional vector algebra!

*Proof B.* Since the two sides of (2) are quadrilinear functions, they must be equal for all $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ once we have checked that their values agree upon some orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. Setting $\mathbf{a} = \mathbf{e}_p, \mathbf{b} = \mathbf{e}_q, \mathbf{c} = \mathbf{e}_r, \mathbf{d} = e_s$ we thus "merely" have to

check that the 81 equations

$$\sum_i \varepsilon_{pqi}\varepsilon_{rsi} = \delta_{pr}\delta_{qs} - \delta_{ps}\delta_{qr} \tag{5}$$

hold for the permutation symbol $\varepsilon_{ijk}$ and the Kronecker delta $\delta_{ij}$.

*Comment.* Carrying out this check is tedious and not very illuminating. Faced with such an unappetizing task, there is a danger that students will develop a life-long aversion to vector algebra (and indeed to tensors). Proof B puts the cart before the horse: surely it is better to prove (2) first, by other means, and then to deduce (5) by taking special values for **a**, **b**, **c**, **d**?

*Proof C.* This assumes a familiarity with the exterior algebra $\wedge E$ of an inner product space $E$, and in particular with the star operator $*: \wedge^p E \to \wedge^{n-p} E$, $n = \dim E$. Using the induced inner product $\langle | \rangle$ on $\wedge^p E$ such that

$$\langle a_1 \wedge \cdots \wedge a_p | b_1 \wedge \cdots \wedge b_p \rangle = \det(a_i \cdot b_j),$$

it is known that $*$ is an isometry. In the case $n = 3$ we define the vector cross product by

$$\mathbf{a} \times \mathbf{b} = *(\mathbf{a} \wedge \mathbf{b}) \tag{6}$$

and (see, e.g., [**9**, p. 419]) derive its usual properties. In particular the isometry property of $*$ yields

$$(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) = \langle \mathbf{a} \wedge \mathbf{b} | \mathbf{c} \wedge \mathbf{d} \rangle,$$

which is the desired identity (2).

*Comments.* Clearly, from the viewpoint of first-time teaching, Proof C is ruled out, but it is well worth mentioning at a more advanced level when teaching exterior algebra. Incidentally one of the virtues of the definition (6), equally well shared by the definition of $\mathbf{a} \times \mathbf{b}$ by (3), is that it makes clear that an $n$-dimensional generalization of the vector cross product exists—namely that vector cross product $a_1 \times a_2 \times \cdots \times a_{n-1}$ of $n-1$ vectors in $n$-dimensions defined by

$$a_1 \times a_2 \times \cdots \times a_{n-1} = *(a_1 \wedge a_2 \wedge \cdots \wedge a_{n-1}).$$

This new product has very many properties that are direct generalizations of familiar 3-dimensional properties (see [**10**]). For a start it satisfies the obvious $n$-dimensional generalization of the axioms (X1)–(X3). It is even possible to deal with the equation

$$a_1 \times \cdots \times a_{n-1} \times v = b$$

on lines directly analogous to the familiar treatment of the equation $\mathbf{a} \times \mathbf{v} = \mathbf{b}$ in three dimensions (again, see [**10**]).

*Proof D.* Both sides of (2) are, for given **a**, **b**, bilinear functions of **c**, **d**. So (2) follows once we have checked its validity for **c**, **d** taking values from an orthonormal basis. This is very easily done. Noting the skew symmetry in **c**, **d** there are effectively only three cases to consider. If $\mathbf{c} = \mathbf{e}_1, \mathbf{d} = \mathbf{e}_2$, for example, then (2) asserts that

$$(\mathbf{a} \times \mathbf{b})_3 = a_1 b_2 - a_2 b_1,$$

which is the well-known result for the third component of $\mathbf{a} \times \mathbf{b}$.

*Comments.* This proof certainly has the merit of brevity. However, it has two less satisfactory aspects. First, it suffers pedagogically from starting out from a position of "knowing the answer." Second, some people may judge it a defect that it makes use of a basis. On the other hand it should be pointed out that these two defects are shared by many other proofs, such as Proof E, which we now outline without further comment.

*Proof E.* We may choose an orthonormal basis such that $\mathbf{a} = a_1\mathbf{e}_1, \mathbf{b} = b_1\mathbf{e}_1 + b_2\mathbf{e}_2, \mathbf{c} = c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + c_3\mathbf{e}_3$. Then, after an easy check, the two sides of (1) are seen to be in agreement.

*Proof F.* On appeal to the geometric definitions of $\mathbf{a} \times \mathbf{b}$ and $\mathbf{a} \cdot \mathbf{b}$ we see that (2) holds in the special case $\mathbf{a} = \mathbf{c}, \mathbf{b} = \mathbf{d}$:

$$(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{a} \times \mathbf{b}) = (\mathbf{a} \cdot \mathbf{a})(\mathbf{b} \cdot \mathbf{b}) - (\mathbf{a} \cdot \mathbf{b})^2 \tag{7}$$

on account of the trigonometric result $\sin^2\theta = 1 - \cos^2\theta$. (Alternatively, in the axiomatic approach, (7) is axiom (X3).) Linearizing (7) in the vector $\mathbf{b}$ we obtain

$$(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{a} \times \mathbf{d}) = (\mathbf{a} \cdot \mathbf{a})(\mathbf{b} \cdot \mathbf{d}) - (\mathbf{a} \cdot \mathbf{d})(\mathbf{a} \cdot \mathbf{b}), \tag{8}$$

after noting that each side of (8) is symmetric under the interchange of $\mathbf{b}$ and $\mathbf{d}$. Consider now the quadrilinear form $\psi$ defined by

$$\psi(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}) = (\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) - (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}) + (\mathbf{a} \cdot \mathbf{d})(\mathbf{b} \cdot \mathbf{c}).$$

Trivially $\psi(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d})$ is zero if $\mathbf{a} = \mathbf{b}$ or if $\mathbf{c} = \mathbf{d}$, but we have just seen in (8) that $\psi(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d})$ is also zero if $\mathbf{a} = \mathbf{c}$. Hence $\psi$ is alternating. But the only alternating function of *four* vectors in *three*-dimensional space is the zero function.

*Comments.* This also is quite a short proof, and it has the further merit of being coordinate-free (except perhaps insofar that a basis is usually employed in the proof of the alternating property to which the proof appeals). From the point of view of first-time teaching the proof is slightly sophisticated and so presumably needs to be delayed until alternating ideas have been encountered in the treatment of determinant functions.

The proof has a further interest. For suppose one pursues the axioms (X1)–(X3) and asks whether they can be satisfied in dimension $n > 3$. Then, as just proved, $\psi$ is alternating. Could it be simply related to another alternating form associated with the axioms (X1)–(X3), namely to the "scalar triple product" $\phi$ defined by $\phi(\mathbf{a}, \mathbf{b}, \mathbf{c}) = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$? One possibility stands out, namely that $\psi$ and $\phi$, viewed as elements of $\wedge^4 E$ and $\wedge^3 E$, are related by $\psi = *\phi$—but take note therefore that the only hope for such a possibility to be realized is in the case $\dim E = 4 + 3 = 7$.

Now it is in fact known that for $n > 3$ axioms (X1)–(X3) can be satisfied *only* in dimension $n = 7$! This surprising result has been shown (see [2, 5]) by appeal to the properties of composition algebras (but see also [11] for a different view). The exceptional nature of dimension 7 in this respect can be related to the existence of the (nonassociative) algebra of octonions in dimension 8, the multiplication of imaginary octonions being given in terms of the dot and cross products in 7 dimensions by

$$\mathbf{a}\,\mathbf{b} = -(\mathbf{a} \cdot \mathbf{b})1 + \mathbf{a} \times \mathbf{b},$$

in direct analogy with the well-known multiplication law for the 3-dimensional imaginary quaternions.

Furthermore it turns out (see [**7**, p. 116]) that the conjectured generalization of the 3-dimensional identity (2), namely

$$(\mathbf{a}\times\mathbf{b})\cdot(\mathbf{c}\times\mathbf{d}) = (\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d}) - (\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{c}) + (*\phi)(\mathbf{a},\mathbf{b},\mathbf{c},\mathbf{d}) \qquad (9)$$

does indeed hold in dimension 7 (for a suitably-signed star operator). What is more surprising still, (9) may conceivably have relevance to our understanding of the fundamental forces of nature! For the identity (9), at least in its coordinate form, has been used in two recent papers [**6**, **13**] that deal with 11-dimensional supergravity theories. (Of the 11 dimensions, the 7 related to the octonions are curled up sufficiently tightly so as to elude direct detection, leaving the familiar 4 (= 3 space + 1 time) dimensions of more conventional theories.) For a coordinate-free proof of the 7-dimensional identity (9) consult [**12**], where it is in fact derived from an 8-dimensional identity for a *ternary* cross product $a \times b \times c$ (obeying the obvious generalization of axioms (X1)–(X3)).

*Remark.* The identity (8), which holds in 7 dimensions just as in 3 dimensions, entails the special case $\mathbf{c} = \mathbf{a}$ of identity (1):

$$(\mathbf{a}\times\mathbf{b})\times\mathbf{a} = (\mathbf{a}\cdot\mathbf{a})\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\mathbf{a}.$$

In consequence the general solution of the equation

$$\mathbf{a}\times\mathbf{v} = \mathbf{b}, \qquad (\mathbf{a}\cdot\mathbf{b} = 0, \qquad \mathbf{a} \neq \mathbf{0})$$

is seen to be given in 7 dimensions by

$$\mathbf{v} = (\mathbf{a}\cdot\mathbf{a})^{-1}(\mathbf{b}\times\mathbf{a}) + \lambda\mathbf{a}, \qquad (\lambda \in \mathbb{R}),$$

exactly as in 3 dimensions.

*Remark.* If one starts out in dimension $n$ ($\geqslant 2$) with (2), or equivalently with (5), then one can easily prove that $n$ must equal 3. If instead one starts out from the weaker version (8) of (2)—or equivalently from the corresponding weaker version of (5) in which a symmetrization has been applied to the indices $p$, $r$—then it is much harder to show that $n$ must equal 3 or 7.

*Recommendations.* Returning now to the teaching of 3-dimensional vector algebra, we believe that the following is the best approach to the identities (1) and (2). First motivate (1) as in Step 1 of Proof A, and note its equivalence with (2). Then switch to Proof D. At a later stage, when teaching determinant functions, mention Proof F.

**3. Appendix.** We can complete Proof A by appealing to a general result on multilinear functions, which we have not seen explicitly stated, though it may be "folklore." The geometric content of the result is that a scalar-valued multilinear function is determined, up to multiplication by a constant scalar, by its zero set.

THEOREM. *Let $V$ be a vector space over a field $F$. Suppose that $f$, $g$ are $p$-linear functions $V^p \to F$ such that*

$$f(v_1,\dots,v_p) = 0 \textit{ whenever } g(v_1,\dots,v_p) = 0.$$

*Then $f = \lambda g$ for some $\lambda \in F$.*

*Remark.* The statement of the theorem avoids the difficulty mentioned in Proof A, concerning the domain of definition of the function $\lambda$ in equation (4).

*Proof.* Let $H_p$ denote the assertion of the theorem. Then $H_p$ is equivalent to the statement that if $f$, $g$ are $p$-linear and

$$f(v_1,\ldots,v_p)=0 \text{ whenever } g(v_1,\ldots,v_p)=0 \tag{10}$$

then

$$\frac{f(x_1,\ldots,x_p)}{g(x_1,\ldots,x_p)}=\frac{f(y_1,\ldots,y_p)}{g(y_1,\ldots,y_p)} \tag{11}$$

whenever $g(x_1,\ldots,x_p)\neq 0$ and $g(y_1,\ldots,y_p)\neq 0$.

We give a proof by induction on the number $p$ of vector arguments of $f$ and $g$. First we show that $H_1$ is true. Given $g(x_1)\neq 0$ and $g(y_1)\neq 0$, set $z=g(x_1)y_1-g(y_1)x_1$. Then $g(z)=g(x_1)g(y_1)-g(y_1)g(x_1)=0$, whence, by (10), $f(z)=0$, that is, $g(x_1)f(y_1)-g(y_1)f(x_1)=0$, so that the $p=1$ version of (11) indeed holds.

We now show that $H_1$ and $H_p$ together imply $H_{p+1}$. To this end consider two $(p+1)$-linear functions $f$, $g$ such that

$$f(v_0,v)=0 \quad \text{whenever } g(v_0,v)=0. \tag{12}$$

(For convenience we adopt the abbreviated notation $f(v_0,v)$ for $f(v_0,v_1,\ldots,v_p)$, etc.) Given also that

$$g(x_0,x)\neq 0 \quad \text{and} \quad g(y_0,y)\neq 0$$

we have to prove that

$$\frac{f(x_0,x)}{g(x_0,x)}=\frac{f(y_0,y)}{g(y_0,y)}. \tag{13}$$

We split the proof into three cases:

(a) $g(x_0,y)\neq 0$.
(b) $g(y_0,x)\neq 0$.
(c) $g(x_0,y)=0=g(y_0,x)$.

*Case (a).* Consider the $p$-linear functions $f(x_0,\cdot\,)$, $g(x_0,\cdot\,)$ and the linear functions $f(\cdot\,,y)$, $g(\cdot\,,y)$. Then

$$\frac{f(x_0,x)}{g(x_0,x)}=\frac{f(x_0,y)}{g(x_0,y)}=\frac{f(y_0,y)}{g(y_0,y)}$$

upon using first $H_p$ and second $H_1$.

*Case (b).* This goes through similarly, by applying $H_1$ to the linear functions $f(\cdot\,,x)$, $g(\cdot\,,x)$ and $H_p$ to the $p$-linear functions $f(x_0,\cdot\,)$, $g(x_0,\cdot\,)$:

$$\frac{f(x_0,x)}{g(x_0,x)}=\frac{f(y_0,x)}{g(y_0,x)}=\frac{f(y_0,y)}{g(y_0,y)}.$$

*Case (c).* Given that $g(y_0,x)=0$ and $g(x_0,y)=0$ it follows from multilinearity that

$$(0\neq)g(x_0,x)=g(x_0+y_0,x) \quad \text{and} \quad g(x_0+y_0,y)=g(y_0,y)(\neq 0).$$

But by (12) we also have $f(y_0, x) = 0$ and $f(x_0, y) = 0$, and hence

$$f(x_0, x) = f(x_0 + y_0, x) \quad \text{and} \quad f(x_0 + y_0, y) = f(y_0, y).$$

Consequently,

$$\frac{f(x_0, x)}{g(x_0, x)} = \frac{f(x_0 + y_0, x)}{g(x_0 + y_0, x)} = \frac{f(x_0 + y_0, y)}{g(x_0 + y_0, y)} = \frac{f(y_0, y)}{g(y_0, y)},$$

where the second equality results from applying $H_p$ to the $p$-linear functions $f(x_0 + y_0, \cdot\,)$, $g(x_0 + y_0, \cdot\,)$.

Thus we have proved (13) in all three cases, and so the argument by induction on $p$ succeeds.

*Proof* A (resumed). Define quadrilinear functions $f, g : E^4 \to R$ by

$$f(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}) = (\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) (= \{(\mathbf{a} \times \mathbf{b}) \times \mathbf{c}\} \cdot \mathbf{d})$$

and

$$g(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}) = (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}) - (\mathbf{a} \cdot \mathbf{d})(\mathbf{b} \cdot \mathbf{c}).$$

Step 1 of the proof (after forming dot products with the vector $\mathbf{d}$) entails that

$$f(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}) = 0 \text{ whenever } g(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}) = 0.$$

Upon noting that $f(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_1, \mathbf{e}_2)/g(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_1, \mathbf{e}_2) = 1/1 = 1$, the theorem thus yields the desired identity (2):

$$f(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}) = g(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}).$$

REFERENCES

1. R. A. Barnett and J. N. Fujii, Vectors, John Wiley, 1963 (reprint: Robert E. Krieger Pub. Co., Huntington, New York, 1975).
2. R. B. Brown and A. Gray, Vector cross products, *Comment. Math. Helv.*, 42 (1967) 222–236.
3. S. Chapman and E. A. Milne, The proof of the formula for the vector triple product, *Math. Gazette*, 23 (1939) 35–38.
4. J. Dieudonné, Linear Algebra and Geometry, Kershaw, London, 1969.
5. B. Eckmann, Stetige Lösungen Linearer Gleichungssysteme, *Comment. Math. Helv.*, 15 (1942–3) 318–339.
6. F. Gürsey and C-H Tze, Octonionic torsion on $S^7$ and Englert's compactification of $d = 11$ supergravity, *Phys. Lett.*, 127B (1983) 191–196.
7. R. Harvey and H. B. Lawson, Calibrated geometries, *Acta Math*, 148 (1982) 47–157.
8. R. Shaw, The teaching of vector algebra, *Int. J. Math. Educ. Sci. Technol.*, 16 (1985) 593–602.
9. ______, Linear Algebra and Group Representations, vol. 2, Academic Press, London, 1983.
10. ______, Vector cross products in $n$ dimensions, *Int. J. Math. Educ. Sci. Technol.*, 18 (1987) 803–816.
11. ______, Ternary vector cross products, *J. Phys. A: Math. Gen.* 20 (1987) L689–L694.
12. ______, A six-variable identity for a ternary vector cross product in 8-dimensional space, *J. Phys. A: Math. Gen.*, 21 (1988) 593–597.
13. B. de Wit and H. Nicolai, The parallelizing $S^7$ torsion in gauged $N = 8$ supergravity, *Nucl. Phys.*, B231 (1984) 506–532.

has explicit geometric meaning. (4) The entire argument can be summarized by the statement, "Apply the Pythagorean Theorem to the right triangle of the figure."

Our proof is essentially the proof given without motivation or interpretation in [1, p. 61].

REFERENCE

1. Marvin Marcus and Henryk Minc, A Survey of Matrix Theory and Matrix Inequalities, Prindle, Weber & Schmidt, Boston, 1964.

## From Experimentation to Proof

HERVÉ LEHNING
*13 rue Letellier, 75015 Paris, France*

I particularly took notice of the following problem, proposed by A. Tissier in the MONTHLY [1]:

"Prove that the differential equation $y' = x - 1/y$ has a unique solution on $[0, \infty)$ which is positive throughout and tends to zero at $+\infty$."

Before the appearance of personal computers, this kind of problem was hard to assign to students because they found it difficult to see the family of solution curves of a differential equation without solving it by quadratures. In this paper, I would like to use this example to demonstrate an intuitive and experimental approach that not only points to the result but also gives some ideas for the proof.

**1. Research.** The first idea is to sketch the family of curves by using adapted software (see, for example [2]). More precisely, we sketch the solution curves of initial value problems:

$$y' = x - \frac{1}{y}, \qquad y(0) = y_0 > 0.$$

By trying a number of values for $y_0$, we quickly notice that if $y_0$ is too large, the solution does not seem to be bounded and if $y_0$ is too small, it does not seem to be defined in all of $[0, \infty)$. More precisely, our successive trials are: 1, 2, 1.5, 1.2, 1.3, and 1.25. You can see the result in FIGURE 1.

Intuitively, Tissier's result seems correct; moreover the solution seems to be the limit of two sequences; the first one is increasing and the other one is decreasing. More precisely, we can define (in fact, for the time being, it is just a conjecture; we shall have to prove it) two sequences of functions $f_n$ and $g_n$ (see FIGURE 2) that seem to converge to the solution we are looking for.

We also notice that a curve (which is $y = 1/x$ according to the study of the sign of $y' = x - 1/y$) divides the first quadrant into regions where the solution is increasing and decreasing.

See [3, 4, 5] for other examples of this kind of approach.

**2. From research to proof.** In fact, if we prove that these sequences are well defined, it is easy to show that $f_n$ is decreasing and $g_n$ increasing (uniqueness of the solution of the initial value problem); and then: $|f_n - g_n| \leqslant 1/n$ because $f_n - g_n$ is

has explicit geometric meaning. (4) The entire argument can be summarized by the statement, "Apply the Pythagorean Theorem to the right triangle of the figure."

Our proof is essentially the proof given without motivation or interpretation in [1, p. 61].

REFERENCE

1. Marvin Marcus and Henryk Minc, A Survey of Matrix Theory and Matrix Inequalities, Prindle, Weber & Schmidt, Boston, 1964.

## From Experimentation to Proof

HERVÉ LEHNING
*13 rue Letellier, 75015 Paris, France*

I particularly took notice of the following problem, proposed by A. Tissier in the MONTHLY [1]:

"Prove that the differential equation $y' = x - 1/y$ has a unique solution on $[0, \infty)$ which is positive throughout and tends to zero at $+\infty$."

Before the appearance of personal computers, this kind of problem was hard to assign to students because they found it difficult to see the family of solution curves of a differential equation without solving it by quadratures. In this paper, I would like to use this example to demonstrate an intuitive and experimental approach that not only points to the result but also gives some ideas for the proof.

**1. Research.** The first idea is to sketch the family of curves by using adapted software (see, for example [2]). More precisely, we sketch the solution curves of initial value problems:

$$y' = x - \frac{1}{y}, \qquad y(0) = y_0 > 0.$$

By trying a number of values for $y_0$, we quickly notice that if $y_0$ is too large, the solution does not seem to be bounded and if $y_0$ is too small, it does not seem to be defined in all of $[0, \infty)$. More precisely, our successive trials are: 1, 2, 1.5, 1.2, 1.3, and 1.25. You can see the result in FIGURE 1.

Intuitively, Tissier's result seems correct; moreover the solution seems to be the limit of two sequences; the first one is increasing and the other one is decreasing. More precisely, we can define (in fact, for the time being, it is just a conjecture; we shall have to prove it) two sequences of functions $f_n$ and $g_n$ (see FIGURE 2) that seem to converge to the solution we are looking for.

We also notice that a curve (which is $y = 1/x$ according to the study of the sign of $y' = x - 1/y$) divides the first quadrant into regions where the solution is increasing and decreasing.

See [3, 4, 5] for other examples of this kind of approach.

**2. From research to proof.** In fact, if we prove that these sequences are well defined, it is easy to show that $f_n$ is decreasing and $g_n$ increasing (uniqueness of the solution of the initial value problem); and then: $|f_n - g_n| \leqslant 1/n$ because $f_n - g_n$ is

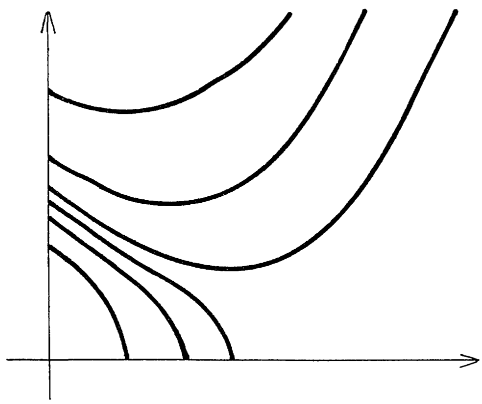


FIG. 1

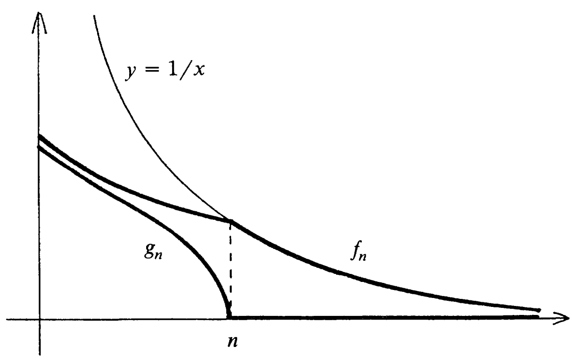



FIG. 2

decreasing in $[0, n]$. We easily deduce that $f_n$ and $g_n$ converge uniformly to the same function $f$. This function is continuous, so by writing the equation in integral form, we can prove that it is a solution of the problem; what follows is easy because if $y$ is a solution of the equation in $[0, \infty)$:

—if $y(0) > f(0)$ then $y(0) > f_n(0)$ for an integer $n$ and so according to the uniqueness theorem, $y > f_n$ for each $x$ so $y$ is not bounded.
—if $y(0) < f(0)$ then, for the same reason, $y$ is not defined in the whole interval.

In order to remain close to the main subject that we would like to stress, we shall not discuss here the search for the proof of the definition of $f_n$ and $g_n$ because it is a classical question. As a matter of fact, it is sufficient to study the solutions of the initial value problem in the open set defined by the inequalities $y > 0$, $xy < 1$, which is easy, as these solutions are necessarily decreasing (nevertheless, please note that in this case too, the experimental approach is useful).

**3. Subtlety.** At this step, we can begin proving. Nevertheless, we see that the correct writing of the proof remains awkward since we must introduce some questions of uniform convergence. The same idea can be used in order to obtain a

more elementary proof. To each $x > 0$, we assign $f(x)$ and $g(x)$ as shown in FIGURE 3.

Then we show that $f$ is strictly decreasing (because there is a unique solution curve passing through each point, see FIGURE 4) and positive, $g$ is strictly increasing and these two functions satisfy

$$0 \leqslant f(x) - g(x) \leqslant 1/x \quad \text{for each } x.$$

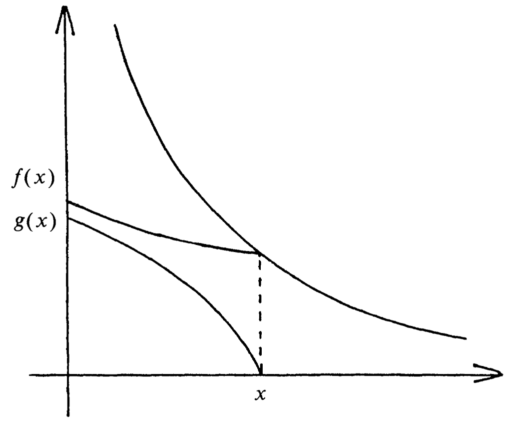



FIG. 3

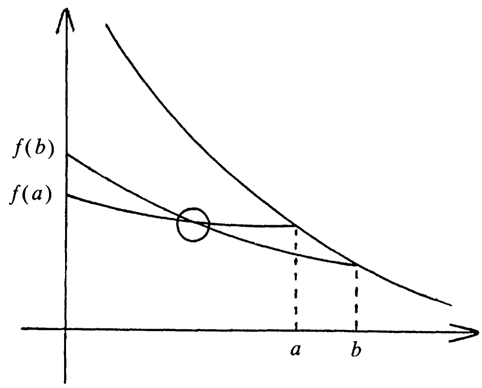



FIG. 4

This allows us to conclude that $f$ and $g$ have a common limit $l$ as $x$ tends to infinity. Then it is easy to show that the solution of the initial value problem at $(0, l)$ is the solution we are looking for. Once these two often hidden, but essential, steps are completed, we can begin to write the proof in its usual baldness. Here, we shall present it in details in order to show that our method does not ruin the notion of rigor.

## 4. Proof

### 4.1. First step: Generalities on the solutions of the equation.

(i) Let $U$ be the open set of the plane defined by the inequalities: $y > 0$, $xy < 1$. The function $(x, y) \mapsto x - 1/y$ is continuously differentiable in $U$ so if $(x_0, y_0)$ belongs to $U$ then the initial value problem:

$$y' = x - \frac{1}{y}, \qquad y(x_0) = y_0, \qquad (x, y) \text{ in } U$$

has a unique maximal solution defined in an open interval $(x_1, x_2)$ where $x_1$ and $x_2$ can be infinite. Because $x - 1/y < 0$, $y$ is decreasing on $(x_1, x_2)$, so: $y \geqslant y_0$ in $(x_1, x_0]$ and then: $y' \geqslant x - 1/y_0$ in $(x_1, x_0]$. Integrating yields

$$y_0 - y \geqslant \frac{1}{2}\left[\left(x_0 - \frac{1}{y_0}\right)^2 - \left(x - \frac{1}{y_0}\right)^2\right], \quad \text{whence}$$

$$y \leqslant y_0 + (x_0 - x)\left[\frac{1}{y_0} - \frac{x + x_0}{2}\right] \text{in } (x_1, x_0].$$

If $x_1$ is finite, we deduce that $y$ has a limit $y_1$ as $x$ tends to $x_1$. The solution $y$ being maximal, $(x_1, y_1)$ does not belong to $U$ so either $y_1 = 0$ or $x_1 y_1 = 1$. But these two cases must be excluded because on the one hand $y_1 \geqslant y_0$ and on the other hand, if $x_1 y_1 = 1$, then $(y - y_1)/(x - x_1)$ tends to 0 as $x$ tends to $x_1$. Now

$$\frac{y - y_1}{x - x_1} \leqslant \frac{\dfrac{1}{x} - \dfrac{1}{x_1}}{x - x_1},$$

which leads to a contradiction when passing to the limit. Then $x_1 = -\infty$.

In the same way, as $x$ tends to $x_2$, $y$ has a finite limit $y_2$ since it is decreasing and positive. If $x_2$ is infinite then $y_2 = 0$ since $xy < 1$; if $x_2$ is finite then either $y_2 = 0$ or $x_2 y_2 = 1$.

So, we have proved that if $(x_0, y_0)$ belongs to $U$ then the initial value problem

$$y' = x - \frac{1}{y}, \qquad y(x_0) = y_0, \qquad (x, y) \text{ in } U$$

has a unique maximal solution $y$ defined in an open interval $(-\infty, a)$ where $a$ is finite or infinite and $y$ tends to $1/a$ or to 0 as $x$ tends to $a$.

(ii) Let $V$ be the open set defined by the inequalities: $y > 0$, $xy > 1$, and $(x_0, y_0)$ belong to $V$. In the same way, the initial value problem

$$y' = x - \frac{1}{y}, \qquad y(x_0) = y_0, \quad (x, y) \text{ in } V$$

has a unique maximal solution $y$ defined in an open interval $(a, \infty)$ where $a > 0$; moreover, $y$ tends to $1/a$ as $x$ tends to $a$ and to $\infty$ as $x$ tends to $\infty$.

(iii) For $a > 0$, the initial value problem

$$y' = x - \frac{1}{y}, \qquad y(a) = \frac{1}{a}, \qquad y > 0$$

has a unique maximal solution $y$. The derivative at point $a$ of the function $x \mapsto xy(x)$ is $1/a > 0$ so there are points $x_0 < a$ and $x_1 > a$ that belong to the interval of definition of $y$ and such that $(x_0, y(x_0))$ belongs to $U$ and $(x_1, y(x_1))$ to $V$. Applying the results obtained in (i) and (ii) at these points, we show that $y$ is defined in $(-\infty, \infty)$.

(iv) So let $W$ be the open set defined by the inequality: $y > 0$. If $(x_0, y_0)$ belongs to $W$, then the initial value problem $y' = x - 1/y$, $y(x_0) = y_0$, $y > 0$ has a unique maximal solution $y$ defined in $(-\infty, a)$ where $a$ is finite or infinite. If $a$ is finite or if $a$ is infinite and the graph does not intersect the curve whose equation is $xy = 1$, the limit of $y$ at $a$ is 0; otherwise it is $+\infty$.

If $y_{1_0} < y_{2_0}$ (for the same $x_0$) then the two associated solutions verify $y_1 < y_2$ in the common interval of definition; otherwise, according to the intermediate value theorem, there would be an $x$ such that $y_1(x) = y_2(x)$ and so $y_1 = y_2$ according to the uniqueness theorem, which is a contradiction.

**4.2. Second step: definition of functions $f$ and $g$.** Let $a > 0$ and $y_a$ the solution in $(-\infty, \infty)$ of the initial value problem $y' = x - 1/y$, $y(a) = 1/a$, $y > 0$. We define $f(a) = y_a(0)$. Then $f$ maps $(0, \infty)$ into itself. Let $a < b$. Then $y_b(a) < y_a(a) = 1/a$ since $(a, y_b(a))$ belongs to $U$. Hence $y_b < y_a$ so $f(a) > f(b)$. We

deduce that $f$ is strictly decreasing. In the same way, in the open set defined by $xy - 1 < 0$, the initial value problem

$$x' = \frac{y}{xy - 1}, \qquad x(0) = a$$

has a unique maximal solution $x_a$ defined in an interval containing $y_0 > 0$. The derivative $x_a'$ does not vanish anywhere in $(0, y_0)$ because $y/(xy - 1) < 0$. Thus, $x_a$ has an inverse function $z$ which is the solution of the equation $y' = x - 1/y$ in $[x_a(y_0), a)$. According to the uniqueness theorem, the maximal solution $z_a$ of the initial value problem $y' = x - 1/y$, $y(x_a(0)) = y_0$, $(x, y)$ in $U$ equals $z$ in this interval. So $z_a$ is defined in $[0, a)$ and tends to 0 as $x$ tends to $a$. Let $g(a) = z_a(0)$. In the same way, we show that $g$ maps $[0, \infty)$ into itself and is strictly increasing, and that $f(x) > g(x)$ for each $x$. Moreover,

$$y_a' - z_a' = \left(x - \frac{1}{y_a}\right) - \left(x - \frac{1}{z_a}\right) = \frac{y_a - z_a}{y_a z_a}.$$

Then $y_a' - z_a' > 0$ and so $f(a) - g(a) < 1/a$.

**4.3. Third step: conclusion.** The function $f$ is decreasing and positive, so it converges to a limit $l$. Let $y$ be the maximal solution of the initial value problem

$$y' = x - \frac{1}{y}, \qquad y(0) = l, \quad y > 0;$$

$y$ is defined in $(-\infty, \infty)$ since otherwise it would be one of the above functions $z_a$ (see first step in (iv)). In the same way, for each $x$, $(x, y(x))$ belongs to $U$, since otherwise it would be one of the functions $y_a$, so $y$ is positive and tends to zero as $x$ tends to $\infty$. It is the unique solution of our problem because if $y$ is another one, let $b = y(0)$.

If $b > l$, then there is an $a$ such that $b > f(a)$ and then $y > y_a$, so it is not bounded.

If $b < l$, then there is an $a$ such that $b < g(a)$ and then $y < z_a$, so $y$ is not defined beyond $a$.

**5. Postscript.** I described this small example in such detail because I think we do not distinguish enough between research and proofs in our teaching. Of course, we all know that mathematical activity requires an experimental phase. But, too often, we imitate Bourbaki's way of writing [**6**], which hides the research and the trial phases completely. So, some students see mathematics as a dead science whereas mathematics is very much alive.

REFERENCES

1. A. Tissier, problem 6551, this MONTHLY, 94 (1987) 694.
2. Atelier Logiciel de l'Ecole Centrale, Etude des suites, Fil, Paris, 1987.
3. Michèle Artigue, Systèmes Différentiels, Cedic, Paris, 1983.
4. Hervé Lehning, Computers as mathematical problem solving assistants, in A Computer for Each Student, R. Lewis and E. D. Tagg, eds., Elsevier/North Holland, Amsterdam, 1987.
5. ______, Analyse Fonctionnelle, Masson, Paris, 1988.
6. Nicolas Bourbaki, Eléments de Mathématiques, Masson, Paris, 1939–1988.

E 3340. *Proposed by Courtney Moen, U. S. Naval Academy, Annapolis, Maryland.*

Suppose $k$ is a given integer $\geqslant 2$. For $n \in \mathbb{N}$ (the set of positive integers) put

$$f(n) = \left\lfloor \left(n + n^{1/k}\right)^{1/k} \right\rfloor + n,$$

where $\lfloor\ \rfloor$ denotes the greatest integer function. What is the range of the function $f$?

E 3341. *Proposed by P. Gritzmann, University of Siegen, West Germany; B. Mohar, University of Ljubljana Yugoslavia; J. Pach, Courant Institute, New York University and Mathematical Institute of the Hungarian Academy of Sciences; and R. Pollack, Courant Institute, New York University.*

Let $G$ be a planar graph obtained from a convex $n$-gon by triangulating its interior, i.e., by subdividing the interior into triangles using non-crossing diagonals. Let $P$ be an arbitrary set of $n$ points in the plane no three of which lie on a line.

Show that $G$ can be straight-line embedded on $P$. I.e., show that there exists a bijection $f$ from the vertex set of $G$ to $P$ such that if $(a, b)$ and $(c, d)$ are distinct edges of $G$, then the open segments $(f(a), f(b))$ and $(f(c), f(d))$ are disjoint.

E 3342. *Proposed by Irving Adler, North Bennington, Vermont.*

Let $S = \{1, 2, \ldots, n^2\}$ and define a permutation $f: S \to S$ as follows. Take $n^2$ cards numbered from 1 to $n^2$ and lay them in a square array, with the $i$th row containing cards $(i-1)n+1, (i-1)n+2, \ldots, in$ in order from left to right. Pick up the cards along rising diagonals, starting with the upper left-hand corner. If card $j$ is the $k$th card picked up, put $f(j) = k$. For example, if $n = 3$, then $f(1) = 1$, $f(4) = 2$, $f(2) = 3$, $f(7) = 4$, $f(5) = 5$, $f(3) = 6$, $f(8) = 7$, $f(6) = 8$, $f(9) = 9$.

For each value of $n$ both 1 and $n^2$ are fixed points of $f$. If $n$ is odd, then $(n^2+1)/2$ is also a fixed point of $f$. Characterize those $n$ for which $f$ has a fixed point not in $\{1, n^2, (n^2+1)/2\}$.

## SOLUTIONS OF ELEMENTARY PROBLEMS

### One Tough Area Problem

E 2983 [1983, 54]. *Proposed by E. Ehrhart, University of Strasbourg, France.*

Let $ABC$ be an equilateral triangle of perimeter $3a$. Calculate the area of the convex region consisting of all points $P$ such that $PA + PB + PC \leqslant 2a$.

*Solution by Greg Fee, University of Waterloo, Ontario, Canada.* Let

$$R = \{P : PA + PB + PC \leqslant 2a\}.$$

Then we shall show that

$$\begin{aligned} a^{-2}\operatorname{area}(R) &= 62\pi/9 - (124/9)\arcsin\left(5\sqrt{3}/9\right) - 20\sqrt{2}/9 \\ &= 0.65504842837277879\ldots. \end{aligned}$$

E 3340. *Proposed by Courtney Moen, U. S. Naval Academy, Annapolis, Maryland.*

Suppose $k$ is a given integer $\geqslant 2$. For $n \in \mathbb{N}$ (the set of positive integers) put

$$f(n) = \left\lfloor \left(n + n^{1/k}\right)^{1/k} \right\rfloor + n,$$

where $\lfloor\ \rfloor$ denotes the greatest integer function. What is the range of the function $f$?

E 3341. *Proposed by P. Gritzmann, University of Siegen, West Germany; B. Mohar, University of Ljubljana Yugoslavia; J. Pach, Courant Institute, New York University and Mathematical Institute of the Hungarian Academy of Sciences; and R. Pollack, Courant Institute, New York University.*

Let $G$ be a planar graph obtained from a convex $n$-gon by triangulating its interior, i.e., by subdividing the interior into triangles using non-crossing diagonals. Let $P$ be an arbitrary set of $n$ points in the plane no three of which lie on a line.

Show that $G$ can be straight-line embedded on $P$. I.e., show that there exists a bijection $f$ from the vertex set of $G$ to $P$ such that if $(a, b)$ and $(c, d)$ are distinct edges of $G$, then the open segments $(f(a), f(b))$ and $(f(c), f(d))$ are disjoint.

E 3342. *Proposed by Irving Adler, North Bennington, Vermont.*

Let $S = \{1, 2, \dots, n^2\}$ and define a permutation $f: S \to S$ as follows. Take $n^2$ cards numbered from 1 to $n^2$ and lay them in a square array, with the $i$th row containing cards $(i-1)n+1, (i-1)n+2, \dots, i\,n$ in order from left to right. Pick up the cards along rising diagonals, starting with the upper left-hand corner. If card $j$ is the $k$th card picked up, put $f(j) = k$. For example, if $n = 3$, then $f(1) = 1$, $f(4) = 2$, $f(2) = 3$, $f(7) = 4$, $f(5) = 5$, $f(3) = 6$, $f(8) = 7$, $f(6) = 8$, $f(9) = 9$.

For each value of $n$ both 1 and $n^2$ are fixed points of $f$. If $n$ is odd, then $(n^2+1)/2$ is also a fixed point of $f$. Characterize those $n$ for which $f$ has a fixed point not in $\{1, n^2, (n^2+1)/2\}$.

## SOLUTIONS OF ELEMENTARY PROBLEMS

### One Tough Area Problem

E 2983 [1983, 54]. *Proposed by E. Ehrhart, University of Strasbourg, France.*

Let $ABC$ be an equilateral triangle of perimeter $3a$. Calculate the area of the convex region consisting of all points $P$ such that $PA + PB + PC \leqslant 2a$.

*Solution by Greg Fee, University of Waterloo, Ontario, Canada.* Let

$$R = \{P: PA + PB + PC \leqslant 2a\}.$$

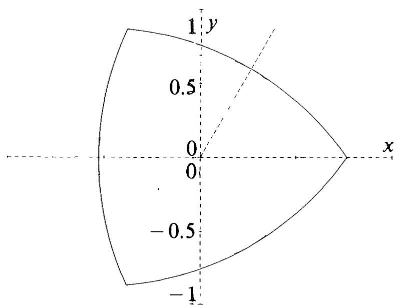
Then we shall show that

$$\begin{aligned} a^{-2}\operatorname{area}(R) &= 62\pi/9 - (124/9)\arcsin\left(5\sqrt{3}/9\right) - 20\sqrt{2}/9 \\ &= 0.65504842837277879\dots. \end{aligned}$$

The algebraic calculations which are needed in what follows were performed on the computer algebra system Maple at the Symbolic Computation Group of the University of Waterloo, although they could certainly be forced through by hand.

By rescaling and shifting we may take $a = \sqrt{3}$, $A = (1, 0)$, $B = (-1/2, \sqrt{3}/2)$, and $C = (-1/2, -\sqrt{3}/2)$. We remark that $R$ is indeed convex: If $P$ and $P'$ lie in $R$ and if $Q = \alpha P + (1 - \alpha)P'$ with $0 < \alpha < 1$, then $Q \in R$ by the triangle inequality.

The triangle $ABC \subset R$, since each of the points $A, B, C \in R$ and $R$ is convex. Also, $R$ is contained within the constant diameter "triangle" created by the intersection of circles centered at $A$, $B$, and $C$, each of radius $\sqrt{3}$. It is easy to see the constant diameter "triangle" is contained in the unit disc centered at the origin, and thus $R$ lies in the unit disc. See the FIGURE.




Let $P = (x, y)$ denote a point on $\partial R$. Expanding the equation $PA + PB + PC = 2\sqrt{3}$ to eliminate radicals, we obtain the polynomial equation

$$3x^8 + 3y^8 + 12x^6y^2 + 18x^4y^4 + 12x^2y^6 - 124(x^6 + y^6) - 372(x^4y^2 + x^2y^4) + 1314(x^4 + y^4) + 2628x^2y^2 + 512x^3 - 1536xy^2 - 3036x^2 - 3036y^2 + 1331 = 0.$$

Converting this to polar coordinates we obtain the surprisingly simple equation

$$512r^3\cos(3\theta) = -3r^8 + 124r^6 - 1314r^4 + 3036r^2 - 1331. \tag{1}$$

This equation exhibits the 120° rotational symmetry of $R$ and symmetry under reflection about the $x$-axis, the 60° line, and the 120° line. Also, we remark that the squaring operations introduced many spurious solutions. For example, when $\theta = 0$, besides the expected double zero $r = 1$, there are also double zeros at $r = 1 \pm 2\sqrt{3}$ and simple zeros at $r = -3 \pm 4/\sqrt{3}$.

Since $A, B, C \in R$ and $R$ lies in the unit disc, $M := \max\{r : re^{i\theta} \in R\} = 1$. Now we show that $m := \min\{r : re^{i\theta} \in \partial R\}$ is achieved at $\theta = \pi/3$, $\pi$, and $5\pi/3$. If we divide through (1) by $r^3$ and differentiate we obtain

$$-512\sin(3\theta) = r^{-4}f(r)\frac{dr}{d\theta}, \tag{2}$$

where

$$f(r) = -5r^8 + 124r^6 - 438r^4 - 1012r^2 + 1331.$$

If we substitute this expression in $I$ and let $x = r^2$, we get

$$I = \frac{-1}{2}\int_{m^2}^{M^2} \frac{5x + 11\,dx}{\{-121 + 258x - 9x^2\}^{1/2}}.$$

This is a standard integral; it equals

$$\left[\frac{5}{18}\{-121 + 258x - 9x^2\}^{1/2} - \frac{124}{9}\arcsin\left\{\frac{18x - 258}{144\sqrt{3}}\right\}\right]_{m^2}^{M^2}$$

$$= \frac{20\sqrt{2}}{9} + \frac{124}{9}\arcsin\big(5/(3\sqrt{3})\big) - 62\pi/9.$$

Thus,

$$|S| = -\frac{\pi m^2}{6} - \frac{10\sqrt{2}}{9} - \frac{62}{9}\arcsin\frac{5}{3\sqrt{3}} + \frac{31\pi}{9},$$

and the area of the scaled figure $R$ is

$$\pi m^2 + 6|S| = \frac{62\pi}{3} - \frac{20\sqrt{2}}{3} - \frac{124}{3}\arcsin\left(\frac{5\sqrt{3}}{9}\right).$$

Finally, we multiply by $a^2/3$ to get the claimed area of the general figure.

No correct solutions were received before the customary deadline. The above solution was prepared by Mr. Fee at the request of the editors.

### A Relative Version of Turán's Theorem

E 3171 [1986, 732]. *Proposed by Dean S. Clark, University of Rhode Island, Kingston.*

In a town where it would require no more than $k$ times as many introductions as there are people to acquaint everyone in the town with everyone else, at least $x\%$ of the population belongs to a certain clique of mutual acquaintances. For a given positive $k$, find the minimum possible value of $x = x(k)$.

*Solution by the proposer.* Let $f(t) = 2(t - k)/(t^2 + t)$. If $k$ is rational, the answer is $x(k) = 100f(\lfloor 2k + 1\rfloor)$. If $k$ is irrational, then $100f(\lfloor 2k + 1\rfloor)$ is an unattainable infimum for $x(k)$. Note that $f(2k) = 1/(2k + 1) = f(2k + 1)$, so the expression for the answer reduces to $100/(2k + 1)$ when $2k + 1$ is an integer. Note also that, using calculus, one finds a unique maximum for $f(t)$ at $t = k + \sqrt{k(k + 1)}$, which is between $2k$ and $2k + 1$.

Suppose the town has $n$ people, and consider the graph whose edges are the pairs of acquainted people. Let $x_1$ be the maximum size of a clique. To show $x \geqslant 100f(\lfloor 2k + 1\rfloor)$, it suffices to show $x_1 \geqslant f(\lfloor 2k + 1\rfloor)n$. We establish a bound by counting non-edges.

For any maximal clique $X$, each person not in $X$ has a non-acquaintance in $X$. Hence the members of a maximum-sized clique account for at least $n - x_1$ of the

If we substitute this expression in $I$ and let $x = r^2$, we get

$$I = \frac{-1}{2}\int_{m^2}^{M^2} \frac{5x + 11\,dx}{\{-121 + 258x - 9x^2\}^{1/2}}.$$

This is a standard integral; it equals

$$\left[\frac{5}{18}\{-121 + 258x - 9x^2\}^{1/2} - \frac{124}{9}\arcsin\left\{\frac{18x - 258}{144\sqrt{3}}\right\}\right]_{m^2}^{M^2}$$

$$= \frac{20\sqrt{2}}{9} + \frac{124}{9}\arcsin\left(5/(3\sqrt{3})\right) - 62\pi/9.$$

Thus,

$$|S| = -\frac{\pi m^2}{6} - \frac{10\sqrt{2}}{9} - \frac{62}{9}\arcsin\frac{5}{3\sqrt{3}} + \frac{31\pi}{9},$$

and the area of the scaled figure $R$ is

$$\pi m^2 + 6|S| = \frac{62\pi}{3} - \frac{20\sqrt{2}}{3} - \frac{124}{3}\arcsin\left(\frac{5\sqrt{3}}{9}\right).$$

Finally, we multiply by $a^2/3$ to get the claimed area of the general figure.

No correct solutions were received before the customary deadline. The above solution was prepared by Mr. Fee at the request of the editors.

### A Relative Version of Turán's Theorem

E 3171 [1986, 732]. *Proposed by Dean S. Clark, University of Rhode Island, Kingston.*

In a town where it would require no more than $k$ times as many introductions as there are people to acquaint everyone in the town with everyone else, at least $x\%$ of the population belongs to a certain clique of mutual acquaintances. For a given positive $k$, find the minimum possible value of $x = x(k)$.

*Solution by the proposer.* Let $f(t) = 2(t - k)/(t^2 + t)$. If $k$ is rational, the answer is $x(k) = 100f(\lfloor 2k + 1\rfloor)$. If $k$ is irrational, then $100f(\lfloor 2k + 1\rfloor)$ is an unattainable infimum for $x(k)$. Note that $f(2k) = 1/(2k + 1) = f(2k + 1)$, so the expression for the answer reduces to $100/(2k + 1)$ when $2k + 1$ is an integer. Note also that, using calculus, one finds a unique maximum for $f(t)$ at $t = k + \sqrt{k(k + 1)}$, which is between $2k$ and $2k + 1$.

Suppose the town has $n$ people, and consider the graph whose edges are the pairs of acquainted people. Let $x_1$ be the maximum size of a clique. To show $x \geqslant 100f(\lfloor 2k + 1\rfloor)$, it suffices to show $x_1 \geqslant f(\lfloor 2k + 1\rfloor)n$. We establish a bound by counting non-edges.

For any maximal clique $X$, each person not in $X$ has a non-acquaintance in $X$. Hence the members of a maximum-sized clique account for at least $n - x_1$ of the

decreasing function of $r$ for fixed $n$. Thus the clique size that is forced is

$$r(k,n)=\min\left\{r: qs+\binom{q}{2}r\leqslant kn\right\}.$$

For the percentage of the population that must belong to a single clique, we have $x(k,n)=100r(k,n)/n$, and $x(k)=\inf_n x(k,n)$. If $qs+\binom{q}{2}r=kn$ for some $r$ dividing $n$, then the result of setting $q=n/r$ and $s=0$ is $r=n/(2k+1)$, or $x(k,n)=100/(2k+1)$. However, the specified equality holds if and only if $2k+1$ is an integer, and it is difficult to use this approach to determine the function $x(k,n)$ or $x(k)$ exactly when $2k+1$ is not an integer.

Partially solved by R. H. Jeurissen (The Netherlands).

### A Circular Concurrence

E 3177 [1986, 811]. *Proposed by Jordi Dou, Barcelona, Spain.*

Let $A, B, C$ be three points on a circle. Let $A_1$ (respectively, $B_1, C_1$) be the intersection of the tangent line at $A$ (respectively, $B, C$) with the line through $BC$ (respectively, $CA$, $AB$).

Prove that the circles $ABB_1$, $BCC_1$, $CAA_1$ and the line $A_1B_1C_1$ have a common point.

*Solution I by Helen M. Marston (retired), Douglass College, New Brunswick, NJ.* Avoiding degeneracies, we assume that triangle $ABC$ has angles $\alpha>\beta>\gamma$. Since $\overline{AA_1}^2=\overline{A_1C}\cdot\overline{A_1B}$, triangles $AA_1C$ and $BA_1A$ are similar, as are the pairs $BB_1A, CB_1B$ and $CC_1B, AC_1C$. Using this and $\alpha+\beta+\gamma=\pi$, we find that the angles made by the lines of the triangle and the tangent at a vertex are $\alpha, \beta, \gamma, \alpha, \beta, \gamma$ in cyclic order.

Let $R$ be the other intersection of circle $ABB_1$ with circle $CAA_1$. We use the fact that the measure of an inscribed angle is always half the measure of the subtended arc to compute additional angles. With the assumption $\alpha>\beta>\gamma$, we have $\angle ABB_1=\angle ACA_1=\gamma$ (otherwise one or both might be $\pi-\gamma$), which yields $\angle ARB_1=\gamma$ and $\angle ARA_1=\pi-\gamma$. Hence, $A_1$, $R$, $B_1$ are collinear, with $R$ between $A_1$ and $B_1$.

Similarly, let $R'$ be the intersection of circle $CAA_1$ with circle $BCC_1$. With $\alpha>\beta>\gamma$, we have $\angle CAA_1=\pi-\beta$ and $\angle CBC_1=\beta$, yielding $\angle CR'A_1=\pi-\beta$ and $\angle CR'C_1=\beta$. Hence $A_1, R', C_1$ are collinear, with $R'$ between $A_1$ and $C_1$.

We show $R=R'$ by determining the angles $A_1RB$ and $A_1R'B$. Since $\angle B_1AB=\pi-\alpha$, we have $\angle B_1RB=\alpha$ and $\angle A_1RB=\pi-\alpha$. Similarly $\angle C_1CB=\pi-\alpha$ implies $\angle C_1R'B=\alpha$ and $\angle A_1R'B=\pi-\alpha$. In order for $A_1RB$ and $A_1R'B$ to have equal measure, either $A_1$, $R$, $R'$, $B$ lie on a circle (impossible by construction) or $R=R'$.

*Solution II by O. P. Lossers, Eindhoven University of Technology, Eindhoven, The Netherlands.* We put $\overline{AA_1}=R_a, \overline{BB_1}=R_b, \overline{CC_1}=R_c$. The circles $\Gamma_a:=(A_1,R_a)$, $\Gamma_b:=(B_1,R_b)$, $\Gamma_c:=(C_1,R_c)$ are the Apollonian circles of $\Delta ABC$. The following facts about these circles are well-known: they have two points $U$ and $V$ in common (hence $A_1$, $B_1$, and $C_1$ are on a line $l$) and intersect the circumcircle of $\Delta ABC$ at right angles.

If we assume $a > b > c$ the point $B_1$ is between $A_1$ and $C_1$ and $\angle A_1UB_1 = \angle B_1UC_1 = \pi/3$ (FIG. 1). Suppose that the circle $CAA_1$ meets $l$ at $S$. Then the inner product

$$\left(\overrightarrow{B_1S}, \overrightarrow{B_1A_1}\right) = \left(\overrightarrow{B_1A}, \overrightarrow{B_1C}\right) = R_b^2 = \left(\overrightarrow{B_1B}, \overrightarrow{B_1B}\right). \tag{1}$$

In $\Delta A_1UB_1$ we find: $\angle B_1A_1U < \pi/3 = \angle B_1UA_1$. Hence $\overline{B_1A_1} > \overline{B_1B}$ and therefore $\overline{B_1S} < \overline{B_1A_1}$. This means that $S$ is between $B_1$ and $A_1$. We conclude that $\Delta B_1US$ is directly similar to $\Delta B_1A_1U$ (see (1)).

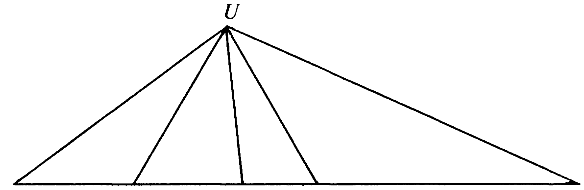



FIG. 1

It follows that $\angle B_1SU = \pi/3$ and $\angle A_1SU = 2\pi/3$. This implies that $\Delta C_1US$ is directly similar to $\Delta C_1B_1U$ and that $\Delta A_1US$ is directly similar to $\Delta A_1C_1U$. Therefore $(\overrightarrow{C_1S}, \overrightarrow{C_1B_1}) = R_b^2 = (\overrightarrow{C_1B}, \overrightarrow{C_1A})$ and $(\overrightarrow{A_1S}, \overrightarrow{A_1C_1}) = R_c^2 = (\overrightarrow{A_1C}, \overrightarrow{A_1B})$. It follows that the circumcircles of the triangles $ABB_1$ and $BCC_1$ pass through $S$. This proves the theorem.

We remark that if $T$ is the point on $l$ such that $\Delta STU$ is equilateral, then it is easily seen that the circles $BAA_1$, $CBB_1$, $ACC_1$ have the point $T$ in common.

*Editorial comment.* Daniel Hurwitz gave a proof using analytic geometry. Several solvers omitted the needed proof that $A_1$, $B_1$, $C_1$ are collinear. Jiro Fukuta suggested the following variant: Let $A, B, C, D$ be four points on a circle. Let the lines $BC$ and $DA$ meet at $E$, $AB$ and $CD$ meet at $F$, the tangents at $A$ and $C$ meet at $G$, and the tangents at $B$ and $D$ meet at $H$. Prove that the circles $ABE$, $CDE$, $BCF$, $DAF$, $ACG$, and $BDH$ have a common point on the line $EFGH$.

Also solved by J. Fukuta (Japan), H. Guggenheimer, J. Heuver (Canada), P. L. Hon (Hong Kong), R. D. Hurwitz, W. Janous (Austria), F. Pouryoussefi (student, Iran), and the proposer.

### A Triangular Triple Product

E 3199 [1987, 301]. *Proposed by H. Guelicher, Muenster, West Germany.*

In the triangle $ABC$, point $Q$ is on the ray $\overrightarrow{BA}$, point $R$ is on the ray $\overrightarrow{CB}$, and $BQ = CR = AC$. A line parallel to $\overleftrightarrow{AC}$ through $R$ intersects $\overleftrightarrow{CQ}$ in a point $T$. A line parallel to $\overleftrightarrow{BC}$ through $T$ intersects $\overleftrightarrow{AC}$ in a point $S$. Show that:

$$(AC)^3 = AQ \cdot BC \cdot CS.$$

*Solution by Ricardo Perez Marco (student), Barcelona, Spain.* We use the notation $(XYZ)$ for the area of the triangle $XYZ$. We have

$$\frac{AC}{AQ} = \frac{QB}{AQ} = \frac{(BQT)}{(QAT)} = \frac{(BCQ)}{(QCA)} = \frac{(BQT)+(BCQ)}{(QAT)+(QCA)} = \frac{(BCT)}{(ATC)}$$

$$\frac{AC}{CS} = \frac{CA}{CS} = \frac{(ATC)}{(STC)}$$

$$\frac{AC}{BC} = \frac{RC}{BC} = \frac{(CTR)}{(CTB)} = \frac{(STC)}{(BCT)},$$

the last since $STRC$ is a parallelogram. Multiplication of the equations gives $AC^3 = AQ \cdot BC \cdot CS$.

Also solved by S. Arslanagić (Yugoslavia), H. D'Souza, R. Dybvik (Norway), J. Ferrer (Spain), P. L. Hon (Hong Kong), D. C. Khatri, O. P. Lossers (The Netherlands), R. B. Nelsen, F. Pouryoussefi (Iran), I. A. Sakmar (Turkey), M. A. Shayib (Kuwait), J. H. Steelman, H.-F. Yeung (Australia), Zhu H.-L. (China), University of South Alabama Problem Group, and the proposer. Most solvers used similar triangles and the law of sines.

### Incomparable Summands with Small Prime Factors

E 3248 [1988, 51]. *Proposed by Paul Erdös, Hungarian Academy of Sciences, and John Selfridge, Northern Illinois University, De Kalb.*

(a) Let $S$ be the set of positive integers with no prime factors bigger than 3. Prove that every positive integer is expressible as a sum of distinct elements of $S$ such that no summand is a multiple of any other. For example, $19 = 9 + 6 + 4$.

(b) Let $T$ be the set of positive integers with no prime factors other than 2, 5, or 7. Prove that every sufficiently large positive integer is expressible as a sum of distinct elements of $T$ such that no summand is a multiple of any other. For example, $62 = 49 + 8 + 5$.

*Solution by Allan Pedersen, Søborg, Denmark.* For either problem, we say that a decomposition of the type described is *satisfactory*. In each case we apply induction.

(a) Note that $1 = 1, 2 = 2, 3 = 3, 4 = 2^2, 5 = 2 + 3$ are satisfactory. For the inductive step, suppose first that $n = 2m$. Then $m$ has a satisfactory decomposition, and we double each term of this to obtain a satisfactory decomposition of $n$. If $n$ is odd, let $m$ be the highest power of 3 not exceeding $n$, and put $p = (n - m)/2$. We use $m$ as a summand. If $p > 0$, then by the inductive hypothesis $p$ has a satisfactory decomposition $p = \Sigma s_i$. We contend that $n = m + \Sigma(2s_i)$ is satisfactory. We need only verify lack of comparability between $m$ and $2s_i$. Because $m$ is a power of three, it is not a multiple of $2s_i$. If $2s_i$ is a multiple of $m$, then $2s_i \geqslant 2m$, which contradicts $n < 3m$.

(b) We prove that any positive integer other than 3, 6, and 31 has a satisfactory decomposition. The failure for 3 and 6 is immediate; 31 takes a bit longer. It can be verified by a combination of examples and inductive shortcuts that, except for these three numbers, all positive integers through 250 have satisfactory decompositions. For $n > 250$, assume by induction that all positive integers less than $n$ (except the three exceptions) have satisfactory decompositions. If $n$ is even, then doubling each

term of a satisfactory decomposition of $n/2$ yields a satisfactory decomposition of $n$. If $n$ is odd, let $m$ be the largest integer with no prime factors other than 5 and 7 that is less than $n - 62$. Because $175 = 5^2 \cdot 7$, we have $m \geqslant 175$. Since $n - m = 2p$ is an even integer bigger than 62, it has a satisfactory decomposition $\Sigma 2t_i$, where $\Sigma t_i$ is a satisfactory decomposition of $p$. We claim that $m + \Sigma 2t_i$ is a satisfactory decomposition of $n$; since $2t_i$ is even, it does not divide $m$, and we need only show that $2t_i$ is not a multiple of $m$.

If $2t_i$ is a multiple of $m$, an odd number, then $2t_i \geqslant 2m$, and $n \geqslant 3m$. However, this leads to a contradiction with the choice of $m$. Since $m$ is the largest product of 5's and 7's bounded by $n - 62$, we claim $125m/49 \geqslant n - 62$. Since $m \geqslant 175$, the product has at least two 5's or at least two 7's. For the latter case, $125m/49$ is another such product, and for the former case $7m/5 < 125m/49$ is another such product. The maximality of $m$ implies $125m/49 \geqslant n - 62$ in either case. With $n \geqslant 3m$, we have $3m \leqslant 125m/49 + 62$, or $m \leqslant 49 \cdot 62/22 < 150$, contradicting $m \geqslant 175$.

*Editorial comment.* David Wells proved an analogue to (b) when the allowed prime factors are 2, 5, 11 (the exceptional integers are then $\{3, 6, 12, 17, 23, 34\}$) and another analogue when the prime factors are 3, 5, 7 (the largest exception is then 148). J. Surányi showed that if only two prime factors are permitted, then for any pair of factors other than 2 and 3 there are infinitely many exceptions. Lorraine Foster showed that in part (b) every number $n \geqslant 39$ has at least two satisfactory decompositions.

Also solved by C. K. Bailey and R. B. Richter, D. Callan, L. L. Foster, O. P. Lossers (The Netherlands), K. McInturff, J.-M. Monier (France), A. Riese, J. H. Steelman, J. Surányi (Hungary), D. M. Wells, and the proposers. Part (a) only was solved by J. T. Ward and K. E. Lewis.

#### Continuous Functions with Bounded Iterates

E 3252 [1988, 132]. *Proposed by David G. Winslow, Louisiana State University, Baton Rouge, LA.*

(a) Suppose that $h$: $\mathbf{R} \to \mathbf{R}$ is continuous and that $h^n$ is bounded for some positive integer $n$, where $h^2(x) = h(h(x))$, $h^3(x) = h(h(h(x)))$, etc. Prove that either $h$ or $h^2$ is bounded.

(b) Show that for each positive integer $n$ there is a continuous function $h$: $\mathbf{R}^2 \to \mathbf{R}^2$ such that $h^k$ is unbounded for $1 \leqslant k \leqslant n - 1$ but $h^n$ is bounded.

*Solution to (a) by Adam Fieldsteel, Wesleyan University, Middletown, CT.* Suppose to the contrary that $h$: $\mathbf{R} \to \mathbf{R}$ is continuous, $h^n$ is bounded for some $n > 0$, and $h$ is not bounded. Then $h(\mathbf{R})$ is an unbounded connected subset of $\mathbf{R}$. If $h(\mathbf{R}) = \mathbf{R}$, then $h^n(\mathbf{R}) = \mathbf{R}$ for all $n$; hence $h(\mathbf{R})$ is a half-line $L \subset \mathbf{R}$. If $h$ is bounded on $L$, then $h^2$ is bounded. If $h$ is unbounded on $L$, then it is unbounded on every half-line contained in $L$, because it is necessarily bounded on the bounded complement of this half-line in $L$. In particular, $h$ is unbounded on $h(L)$ and, by iteration, on $h^n(L)$ for every $n$, contradicting the assumptions.

*Solution to (b) by Ellen Hertz, National Highway Traffic Safety Administration.* Using polar coordinates, let $h(r, \theta) = (r\sin(n\theta), \theta - 2\pi/n)$ for $2\pi/n \leqslant \theta \leqslant 2\pi$, and $h(r, \theta) = (0, 0)$ otherwise.

*Editorial comment.* Roger Eggleton (Australia) pointed out that $e^{-x}$ is an example of an unbounded $h$ such that $h^2$ is bounded.

Solved by the proposer and 21 others. Two partial solutions were received.

**A Variant of Touchard's Identity**

E 3258 [1988, 259]. *Proposed by Nicolae Gonciulea, Traian College, Drobeta Turnu Severin, Romania.*

Prove that

$$\sum_{j=0}^{n}\binom{n}{j}2^{n-j}\binom{j}{\lfloor j/2\rfloor}=\binom{2n+1}{n},$$

where $\lfloor x\rfloor$ is the largest integer not exceeding $x$.

*Composite solution I, based on solutions by David Callan, University of Bridgeport, CT, Ellen Hertz, National Highway Traffic Safety Administration, Washington, DC, Lou Shapiro, Howard University, Washington, DC, and Louis Thurston, Texas A & I University, Kingsville (independently).* The total number of $n$-element subsets of a $(2n+1)$-element set $S$ is $\binom{2n+1}{n}$. Now suppose that $S$ is partitioned into $n$ pairs and a singleton. An $n$-element subset of $S$ contains exactly one member from some number $n-j$ of pairs. We show that for each $j$ there are $\binom{n}{j}2^{n-j}\binom{j}{\lfloor j/2\rfloor}$ of these subsets. To construct such a subset, we choose the $n-j$ pairs and choose one element from each of them, which can be done in $\binom{n}{j}2^{n-j}$ ways. Next we choose $\lfloor j/2\rfloor$ complete pairs from the $j$ remaining pairs, which can be done in $\binom{j}{\lfloor j/2\rfloor}$ ways, and we take the singleton if $n$ is odd.

The same argument shows that, more generally,

$$\sum_{j=0}^{n}\binom{n}{j}2^{n-j}\binom{j}{\lfloor (m-n+j)/2\rfloor}=\binom{2n+1}{m},$$

where $\binom{j}{i}$ is understood to be zero if $i<0$ or $i>j$.

*Composite solution II, based on the solutions of many readers.* Since the constant term in $(1+x)(x^{-1}+x)^j$ is $\binom{j}{\lfloor j/2\rfloor}$, the sum is equal to the constant term in

$$\begin{aligned}\sum_{j=0}^{n}\binom{n}{j}2^{n-j}(1+x)(x^{-1}+x)^j&=(1+x)(2+x^{-1}+x)^n\\&=(1+x)(2x+1+x^2)^n/x^n\\&=(1+x)^{2n+1}/x^n,\end{aligned}$$

which is $\binom{2n+1}{n}$.

*Editorial comment.* Many solvers observed that the identity in question can be obtained by adding the cases $r=0$ and $r=1$ of the following identity (equivalent to a special case of Vandermonde's identity)

$$\sum_{k}\binom{n}{2k+r}2^{n-2k-r}\binom{2k+r}{k}=\binom{2n}{n-r}, \tag{1}$$

which they proved in various ways. George Andrews noted the paper by numerous authors "On Kemmer's identity in combinatory functions," *Mathematics Student*, 15 (1947) 93–100 devoted to this identity.

Several solvers noted that by combining the terms $j = 2k$ and $j = 2k + 1$, Gonciulea's identity can be reduced to Touchard's identity

$$\sum_k \binom{n}{2k} 2^{n-2k} C_k = C_{n+1},$$

where $C_k$ is the Catalan number $\binom{2k}{k}/(k+1)$. Touchard's identity may be obtained from (1) by setting $r = 1$, replacing $n$ with $n + 1$, and dividing both sides by $n + 1$.

Other extensions of Touchard's identity may be found in H. W. Gould, "Generalization of a formula of Touchard for Catalan numbers," *J. Combinatorial Theory*, (A) 23 (1977) 351–353.

Solved by more than 30 readers and the proposer.

## ADVANCED PROBLEMS

6606. *Proposed by Daniel B. Shapiro, Ohio State University, Columbus.*

Suppose $k$ is a field containing an element $i$ with $i^2 = -1$ and suppose $\tilde{k}$ is the algebraic closure of $k$. Suppose $F$ and $K$ are fields with $k \subseteq F \subseteq K$ and $[K : F]$ finite.

(1) If $\tilde{k} \subseteq K$, prove that in fact $\tilde{k} \subseteq F$.

(2) More generally, if $L$ is an algebraically closed field with $k \subseteq L \subseteq K$, does it follow that $L \subseteq F$?

6607. *Proposed by Ira Gessel, Brandeis University, Waltham, MA.*

The Laguerre polynomials are defined by

$$L_n^{(\alpha)}(x) = (-1)^n \sum_{k=0}^{n} (-1)^k \binom{n+\alpha}{k} \frac{x^{n-k}}{(n-k)!},$$

where $\alpha$ is an arbitrary parameter and $n$ is a non-negative integer. Show that for each non-negative integer $k$ there exists a polynomial $q_k$ of degree $\lfloor k/2 \rfloor$ depending only on $k$ such that

$$L_n^{(\alpha)}(x + n + \alpha) = (-1)^n \sum_{k=0}^{n} q_k(n+\alpha) \frac{x^{n-k}}{(n-k)!}.$$

For example $q_0(u) = 1$, $q_1(u) = 0$, $q_2(u) = -u/2$, $q_3(u) = -u/3$, $q_4(u) = u^2/8 - u/4$.

6608. *Proposed by Oliver D. Anderson, University of Western Ontario, London.*

Let $S$ be the set of all points $(x_1, \ldots, x_n)$ in $R^n$ such that

$$x_k = \sum_{j=0}^{n-k} t_j t_{j+k} \Big/ \sum_{j=0}^{n} t_j^2 \quad (k = 1, 2, \ldots, n),$$

where $t_0 = 1$ and $t_1, \ldots, t_n$ are arbitrary real numbers. It is easy to see that $S$ is bounded and contains the origin as an interior point.

(a) When $n = 2$ prove that $S$ is the convex hull of the set consisting of the point $(0, -\frac{1}{2})$ and the points of the ellipse $2x_1^2 + (4x_2 - 1)^2 = 1$.
(b)* Is $S$ convex for all $n$?
(c)* Is $S$ closed for all $n$?

## SOLUTION OF ADVANCED PROBLEMS

### A Continued Fraction for a Quotient of Bessel Functions

6541. *Proposed by Carl W. Helstrom, University of California, San Diego.*

Derive the continued fraction

$$\frac{K_0(x)}{K_1(x)} = \frac{1}{1+}\,\frac{2u}{1+}\,\frac{u}{1+}\,\frac{3u}{1+}\,\frac{3u}{1+}\cdots\frac{(2k+1)u}{1\quad+}\,\frac{(2k+1)u}{1\quad+}\cdots,$$

where $K_0(x)$ and $K_1(x)$ are modified Bessel functions of the second kind and $u = 1/(4x)$.

*Composite solution by Otto G. Ruehr, Michigan Technological University, Houghton, and Richard A. Askey, University of Wisconsin, Madison.* We shall prove that, for $\operatorname{Re}\nu > -1/2$ and $x \in \mathbb{C} \setminus (-\infty, 0]$,

$$\frac{K_\nu(x)}{K_{\nu+1}(x)} = \frac{1}{1+}\,\frac{2(2\nu+1)u}{1\quad+}\,\frac{(1-2\nu)u}{1\quad+}\,\frac{(2\nu+3)u}{1}$$
$$+\frac{(3-2\nu)u}{1\quad+}\,\frac{(2\nu+5)u}{1\quad+}\cdots, \tag{1}$$

where $u = 1/(4x)$. The proposed problem, with no domain of convergence indicated, is the special case $\nu = 0$ of (1).

Let

$$I(a, b; z) = \int_0^\infty \frac{e^{-u} u^{a-1}\,du}{(1+zu)^b},$$

where $\operatorname{Re} a > 0$, $b$ is arbitrary, and $z \in \mathbb{C} \setminus (-\infty, 0]$. Then [**1**, p. 352]

$$\frac{I(a, b; z)}{I(a, b-1; z)} = \frac{1}{1+}\,\frac{az}{1+}\,\frac{bz}{1+}\,\frac{(a+1)z}{1\quad+}\,\frac{(b+1)z}{1\quad+}\cdots.$$

Letting $a = \nu + 1/2$, $b = 1/2 - \nu$, and $z = 2u$, we find that, for $\operatorname{Re}\nu > -1/2$,

$$\frac{I\left(\nu+\frac{1}{2}, -\nu+\frac{1}{2}; 2u\right)}{I\left(\nu+\frac{1}{2}, -\nu-\frac{1}{2}; 2u\right)} = \frac{1}{1+}\,\frac{(2\nu+1)u}{1\quad+}\,\frac{(1-2\nu)u}{1}$$
$$+\frac{(2\nu+3)u}{1\quad+}\,\frac{(3-2\nu)u}{1\quad+}\cdots. \tag{2}$$

(a) When $n = 2$ prove that $S$ is the convex hull of the set consisting of the point $(0, -\frac{1}{2})$ and the points of the ellipse $2x_1^2 + (4x_2 - 1)^2 = 1$.

(b)* Is $S$ convex for all $n$?

(c)* Is $S$ closed for all $n$?

## SOLUTION OF ADVANCED PROBLEMS

### A Continued Fraction for a Quotient of Bessel Functions

6541. *Proposed by Carl W. Helstrom, University of California, San Diego.*

Derive the continued fraction

$$\frac{K_0(x)}{K_1(x)} = \frac{1}{1+}\,\frac{2u}{1+}\,\frac{u}{1+}\,\frac{3u}{1+}\,\frac{3u}{1+}\cdots\frac{(2k+1)u}{1\quad+}\,\frac{(2k+1)u}{1\quad+}\cdots,$$

where $K_0(x)$ and $K_1(x)$ are modified Bessel functions of the second kind and $u = 1/(4x)$.

*Composite solution by Otto G. Ruehr, Michigan Technological University, Houghton, and Richard A. Askey, University of Wisconsin, Madison.* We shall prove that, for $\operatorname{Re} \nu > -1/2$ and $x \in \mathbb{C} \setminus (-\infty, 0]$,

$$\frac{K_\nu(x)}{K_{\nu+1}(x)} = \frac{1}{1+}\,\frac{2(2\nu+1)u}{1\quad+}\,\frac{(1-2\nu)u}{1\quad+}\,\frac{(2\nu+3)u}{1}$$
$$+\frac{(3-2\nu)u}{1\quad+}\,\frac{(2\nu+5)u}{1\quad+}\cdots, \tag{1}$$

where $u = 1/(4x)$. The proposed problem, with no domain of convergence indicated, is the special case $\nu = 0$ of (1).

Let

$$I(a, b; z) = \int_0^\infty \frac{e^{-u}u^{a-1}\,du}{(1+zu)^b},$$

where $\operatorname{Re} a > 0$, $b$ is arbitrary, and $z \in \mathbb{C} \setminus (-\infty, 0]$. Then [**1**, p. 352]

$$\frac{I(a, b; z)}{I(a, b-1; z)} = \frac{1}{1+}\,\frac{az}{1+}\,\frac{bz}{1+}\,\frac{(a+1)z}{1\quad+}\,\frac{(b+1)z}{1\quad+}\cdots.$$

Letting $a = \nu + 1/2$, $b = 1/2 - \nu$, and $z = 2u$, we find that, for $\operatorname{Re} \nu > -1/2$,

$$\frac{I\left(\nu+\frac{1}{2}, -\nu+\frac{1}{2}; 2u\right)}{I\left(\nu+\frac{1}{2}, -\nu-\frac{1}{2}; 2u\right)} = \frac{1}{1+}\,\frac{(2\nu+1)u}{1\quad+}\,\frac{(1-2\nu)u}{1}$$
$$+\frac{(2\nu+3)u}{1\quad+}\,\frac{(3-2\nu)u}{1\quad+}\cdots. \tag{2}$$

REFERENCES

1. H. S. Wall, Analytic Theory of Continued Fractions, Van Nostrand, Toronto, 1948.
2. G. N. Watson, A Treatise on the Theory of Bessel Functions, 2nd ed., Cambridge University Press, Cambridge, 1966.

Also solved by Thomas N. Delmer. A heuristic solution was provided by the proposer.

6550 [1987; 559]. *Proposed by Walther Janous, Ursulinengymnasium, Innsbruck, Austria.*

Let $n$ and $k$ be integers with $0 \leq k \leq n$. Determine the minimum value of

$$\int_0^1 \{P_{n,k}(x)\}^2\,dx,$$

where $P_{n,k}$ runs through the set of polynomials with real coefficients and degree at most $n$ such that the coefficient of $x^k$ is 1.

*Solution by Ian McGee and Cecil Rousseau, University of Waterloo, Canada.* The minimum value is

$$\left[(2k+1)\binom{2k}{k}^2\binom{n+k+1}{2k+1}^2\right]^{-1}$$

Let $[x^k]P(x)$ denote the coefficient of $x^k$ in the polynomial $P$. To minimize $\int_a^b\{P_{n,k}(x)\}^2 w(x)\,dx$ over polynomials $P_{n,k}$ of degree at most $n$ that satisfy $[x^k]P_{n,k}(x) = 1$, form the orthogonal expansion

$$P_{n,k}(x) = \sum_{j=0}^{n} c_j \phi_j(x),$$

where $\{\phi_j(x)\}$ denotes the sequence of orthonormal polynomials on $(a, b)$ with weight function $w$. Our problem is then to minimize $\sum_{j=0}^n c_j^2$ subject to $\sum_{j=0}^n c_j q_j = 1$, where $q_j = [x^k]\phi_j(x)$. By Cauchy's inequality, it follows that the desired minimum is $K = (\sum_{j=0}^n q_j^2)^{-1}$ and that this minimum is obtained when $c_i = Kq_i$ for $i = 0, \ldots, n$. In the present case $w(x) = 1$ and $\phi_j(x) = \sqrt{2j+1}\,P_j(2x-1)$, where $P_j$ denotes the Legendre polynomial of degree $j$. From Rodrigues' formula, we have

$$\phi_j(x) = \frac{\sqrt{2j+1}}{j!}\frac{d^j}{dx^j}\{x(x-1)\}^j,$$

so

$$q_j = \sqrt{2j+1}\,(-1)^{j-k}\binom{2k}{k}\binom{j+k}{2k}.$$

Thus

$$\begin{aligned}\sum_{j=0}^n q_j^2 &= \binom{2k}{k}^2 \sum_{j=k}^n (2j+1)\binom{j+k}{2k}^2 \\ &= \binom{2k}{k}^2 (2k+1) \sum_{j=k}^n \left[\binom{j+k+1}{2k+1}^2 - \binom{j+k}{2k+1}^2\right] \\ &= \binom{2k}{k}^2 (2k+1)\binom{n+k+1}{2k+1}^2, \qquad (1)\end{aligned}$$

REFERENCES

1. H. S. Wall, Analytic Theory of Continued Fractions, Van Nostrand, Toronto, 1948.
2. G. N. Watson, A Treatise on the Theory of Bessel Functions, 2nd ed., Cambridge University Press, Cambridge, 1966.

Also solved by Thomas N. Delmer. A heuristic solution was provided by the proposer.

6550 [1987; 559]. *Proposed by Walther Janous, Ursulinengymnasium, Innsbruck, Austria.*

Let $n$ and $k$ be integers with $0 \le k \le n$. Determine the minimum value of

$$\int_0^1 \{P_{n,k}(x)\}^2\,dx,$$

where $P_{n,k}$ runs through the set of polynomials with real coefficients and degree at most $n$ such that the coefficient of $x^k$ is 1.

*Solution by Ian McGee and Cecil Rousseau, University of Waterloo, Canada.* The minimum value is

$$\left[(2k+1)\binom{2k}{k}^2\binom{n+k+1}{2k+1}^2\right]^{-1}$$

Let $[x^k]P(x)$ denote the coefficient of $x^k$ in the polynomial $P$. To minimize $\int_a^b \{P_{n,k}(x)\}^2 w(x)\,dx$ over polynomials $P_{n,k}$ of degree at most $n$ that satisfy $[x^k]P_{n,k}(x) = 1$, form the orthogonal expansion

$$P_{n,k}(x) = \sum_{j=0}^{n} c_j \phi_j(x),$$

where $\{\phi_j(x)\}$ denotes the sequence of orthonormal polynomials on $(a, b)$ with weight function $w$. Our problem is then to minimize $\sum_{j=0}^n c_j^2$ subject to $\sum_{j=0}^n c_j q_j = 1$, where $q_j = [x^k]\phi_j(x)$. By Cauchy's inequality, it follows that the desired minimum is $K = (\sum_{j=0}^n q_j^2)^{-1}$ and that this minimum is obtained when $c_i = Kq_i$ for $i = 0, \ldots, n$. In the present case $w(x) = 1$ and $\phi_j(x) = \sqrt{2j+1}\,P_j(2x-1)$, where $P_j$ denotes the Legendre polynomial of degree $j$. From Rodrigues' formula, we have

$$\phi_j(x) = \frac{\sqrt{2j+1}}{j!}\frac{d^j}{dx^j}\{x(x-1)\}^j,$$

so

$$q_j = \sqrt{2j+1}\,(-1)^{j-k}\binom{2k}{k}\binom{j+k}{2k}.$$

Thus

$$\begin{aligned}\sum_{j=0}^n q_j^2 &= \binom{2k}{k}^2 \sum_{j=k}^n (2j+1)\binom{j+k}{2k}^2 \\ &= \binom{2k}{k}^2 (2k+1)\sum_{j=k}^n \left[\binom{j+k+1}{2k+1}^2 - \binom{j+k}{2k+1}^2\right] \\ &= \binom{2k}{k}^2 (2k+1)\binom{n+k+1}{2k+1}^2, \qquad (1)\end{aligned}$$

(Denmark), John Henry Steelman, Douglas B. Tyler, and the proposer. One anonymous solution was also received.

6551 [1987, 694]. *Proposed by A. Tissier, Montfermeil, France*

Prove that the differential equation

$$y' = x - \frac{1}{y}$$

has a unique solution on $[0, +\infty)$ which is positive throughout and tends to zero at $+\infty$.

*Solution I by Richard J. Driscoll, Loyola University, Chicago, Illinois.* Set $f = f(x, y) = x - y^{-1}$. Then $f$ and $f_y$ are continuous in the upper half plane, so for any $a > 0$ the differential equation with the initial condition $y(a) = y_a > 0$ has a unique local solution. For the conditions peculiar to the present problem, a key to both existence and uniqueness is provided by the following exponentiated integral formula.

Say $y = g(x)$ and $y = h(x)$ are positive solutions of the differential equation for $x \in I$ where either $I = [0, T]$ or $[0, \infty)$. Since

$$g' - h' = (g - h)/(gh),$$

for any $a \geq 0$ we have

$$\frac{d}{dx}(g(x) - h(x))\exp\left(-\int_a^x (g(t)h(t))^{-1}\,dt\right) = 0$$

and, hence,

$$g(x) - h(x) = (g(a) - h(a))\exp\left(\int_a^x (g(t)h(t))^{-1}\,dt\right)$$

for $x, a \in I$.

Uniqueness is almost immediate, for if $g(x)$ and $h(x)$ are two distinct solutions, without loss of generality $g(a) > h(a)$ for some $a \geqslant 0$ and hence $g(x) - h(x)$ is strictly increasing for $x \geqslant a$. Thus one of them is *not* tending to 0 at infinity.

For existence, we also introduce certain loci, namely,

$$C_1 = \{(x, y) : xy - 1 = 0\}$$

and

$$C_2 = \{(x, y) : y^3 + xy - 1 = 0\}.$$

Here $C_1$ is the locus of points where solutions to the differential equation have $y' = 0$ while $C_2$, which lies between the $x$-axis and $C_1$, is the corresponding locus for $y'' = 0$. These loci cut the first quadrant into three regions $T$, $B$ and $M$ (top, bottom, and middle) in a useful way, since above $C_1$ (in $T$) we have $y' > 0$ and $y'' < 0$, below $C_2$ (in $B$) we have $y' < 0$ and $y'' < 0$, and between $C_1$ and $C_2$ (in $M$) we have $y' < 0$ and $y'' > 0$ (the behavior we are looking for). A solution to the differential equation that passes through a point above $C_1$ is defined on $[0, \infty)$ since the growth rate is positive but bounded by $x + K$ for some constant $K$. A solution that passes through a point below $C_2$ cannot be extended to $[0, \infty)$ since the graph of the solution is decreasing and concave and passes through a point on the $x$-axis.

(Denmark), John Henry Steelman, Douglas B. Tyler, and the proposer. One anonymous solution was also received.

6551 [1987, 694]. *Proposed by A. Tissier, Montfermeil, France*

Prove that the differential equation

$$y' = x - \frac{1}{y}$$

has a unique solution on $[0, +\infty)$ which is positive throughout and tends to zero at $+\infty$.

*Solution I by Richard J. Driscoll, Loyola University, Chicago, Illinois.* Set $f = f(x, y) = x - y^{-1}$. Then $f$ and $f_y$ are continuous in the upper half plane, so for any $a > 0$ the differential equation with the initial condition $y(a) = y_a > 0$ has a unique local solution. For the conditions peculiar to the present problem, a key to both existence and uniqueness is provided by the following exponentiated integral formula.

Say $y = g(x)$ and $y = h(x)$ are positive solutions of the differential equation for $x \in I$ where either $I = [0, T]$ or $[0, \infty)$. Since

$$g' - h' = (g - h)/(gh),$$

for any $a \geq 0$ we have

$$\frac{d}{dx}(g(x) - h(x))\exp\left(-\int_a^x (g(t)h(t))^{-1}\,dt\right) = 0$$

and, hence,

$$g(x) - h(x) = (g(a) - h(a))\exp\left(\int_a^x (g(t)h(t))^{-1}\,dt\right)$$

for $x, a \in I$.

Uniqueness is almost immediate, for if $g(x)$ and $h(x)$ are two distinct solutions, without loss of generality $g(a) > h(a)$ for some $a \geqslant 0$ and hence $g(x) - h(x)$ is strictly increasing for $x \geqslant a$. Thus one of them is *not* tending to 0 at infinity.

For existence, we also introduce certain loci, namely,

$$C_1 = \{(x, y) : xy - 1 = 0\}$$

and

$$C_2 = \{(x, y) : y^3 + xy - 1 = 0\}.$$

Here $C_1$ is the locus of points where solutions to the differential equation have $y' = 0$ while $C_2$, which lies between the $x$-axis and $C_1$, is the corresponding locus for $y'' = 0$. These loci cut the first quadrant into three regions $T$, $B$ and $M$ (top, bottom, and middle) in a useful way, since above $C_1$ (in $T$) we have $y' > 0$ and $y'' < 0$, below $C_2$ (in $B$) we have $y' < 0$ and $y'' < 0$, and between $C_1$ and $C_2$ (in $M$) we have $y' < 0$ and $y'' > 0$ (the behavior we are looking for). A solution to the differential equation that passes through a point above $C_1$ is defined on $[0, \infty)$ since the growth rate is positive but bounded by $x + K$ for some constant $K$. A solution that passes through a point below $C_2$ cannot be extended to $[0, \infty)$ since the graph of the solution is decreasing and concave and passes through a point on the $x$-axis.

Let $y = y(x) = g(x; b)$ denote a solution (and also the corresponding solution curve) of the differential equation satisfying $y(0) = b$, and set

$$S_1 = \{b : b > 0 \text{ and } g(x; b) \text{ enters } T\}$$

and

$$S_2 = \{b : b > 0 \text{ and } g(x; b) \text{ enters } B\}.$$

Clearly these are disjoint nonempty sets. Will a sufficiently slight change in $b$ preserve the property of entering $T$ (resp. of entering $B$)? The solution curve "enters" before (if indeed ever) it touches the $x$-axis, so the exponentiated integral remains finite at least for a while after entry. Now the exponentiated integral formula says $|g(x; b_1) - g(x; b_2)|$ will be arbitrarily small provided that $|g(0; b_1) - g(0; b_2)| = |b_1 - b_2|$ is sufficently small. Hence $S_1$ and $S_2$ are disjoint nonempty *open* sets. Hence some $b_0 > 0$ is *not* in their union, and $y = g(x; b_0)$ is the solution.

*Editorial Comment.* The majority of our solvers used methods more or less similar to the above. (Driscoll and several others also indicated more constructive approaches to the existence problem.) Howard Morris got some results for certain $y' = x^\alpha - y^{-\beta}$ in this manner, while Preben Alsholm (Denmark) extended it to treat a family of differential equations of the form

$$\frac{dy}{dx} = \frac{h(x)(y - f(x))}{y - g(x)}.$$

*Solution II by P. J. Bushell, University of Sussex, Brighton, England.* If a solution exists it is unique. For suppose that $y_1$ and $y_2$ are solutions and $w = (y_1 - y_2)^2$. Then $w' = 2w/(y_1 y_2) \geqslant 0$ and $w$ is nonnegative, nondecreasing, and tends to zero at infinity. Hence it is identically zero, that is $y_1 = y_2$.

The solution we seek is an integral curve of the plane autonomous system

$$\dot{x} = y, \qquad \dot{y} = xy - 1$$

with $(x, y) = (0, \alpha)$ at $t = 0$, where $\alpha > 0$. Plainly, $\ddot{x} = x\dot{x} - 1$ and

$$\dot{x} = \frac{1}{2}x^2 - (t - \alpha),$$

a simple Riccati equation for $x$.

The unique solution of the problem is given in parametric form by

$$x(t) = -2\frac{\dot{z}(t)}{z(t)}, \qquad y(t) = \frac{1}{z(t)^2}\int_t^\infty z(s)\, s$$

where $z(t) = \mathrm{Ai}(\mu(t - \alpha))$. Here $\mathrm{Ai}(x)$ is the Airy function that tends to zero as $s$ tends to infinity, $\mu = 2^{-1/3}$ and $-\mu\alpha$ is the first negative zero of $\mathrm{Ai}'(s)$. (For an account of the Airy functions see §10.4 of M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover, New York, 1970.) To see this, insert the parametric form of $x(t)$ above into the Riccati equation to obtain $\ddot{z} - (1/2)(t - \alpha)z = 0$, an equation that is simply the differentiated form of $\dot{z}^2 - (1/2)(t - \alpha)z^2 = (1/2)\int_t^\infty z(s)^2\, ds$ (since both $\mathrm{Ai}(x)$ and $\mathrm{Ai}'(x)$ tend to 0 at infinity, the constant of integration is zero). But this last equation is easily seen to be $(d/dt)(-2\dot{z}/z) = (1/z)^2\int_t^\infty z^2\, ds$ if we express $\ddot{z}$ in terms of $z$.

The qualitative nature of the phase portrait can be deduced by considering the signs of $\dot{x}$, $\dot{y}$, $y'$, and $y''$. We see that if $-\mu t_0$ is the first negative zero of $\mathrm{Ai}(s)$, then the solution path given by the parametric solution, with $\alpha = t_0 < t < \infty$, is the separatrix between two families of solution paths with very different behaviour.

*Editorial Comment.* The proposer and J. Bethéry were aware of the explicit approach, and even remarked that the unique solution would pass through $(\sigma, \sigma^2/2) = (1.157233\ldots, 0.669594\ldots)$ where $\sigma = (12)^{1/3}\Gamma(2/3)/\Gamma(1/3)$. Greg Fee of the Maple Symbolic Computation Group at the University of Waterloo (Canada) verified this to about 50 significant digits (at the request of a skeptical editor) and also determined $y(0) = 1.2835987104635995\ldots$. to 50 significant digits. Various planar diagrams of solutions and direction fields, some done on computers and some by hand, were provided by Felix Albrecht, P. J. Bushell (U.K.), Greg Fee, R. I. Raichinov-Simeonov (Bulgaria), and Richard C. Swanson. Fee's graph of the solution is given as FIGURE 1.

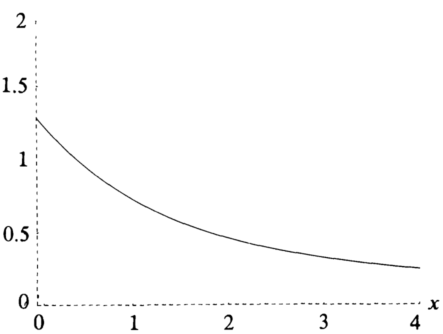



FIG. 1

In particular, David V. V. Wend considered the plane as mapped stereographically onto the Riemann sphere and studied the configuration at $\infty$. He noted that "the image of the separatrix $y(x, y^*)$ [here $y(x, y_0)$ is the solution with $y(0) = y_0$] starts out at infinity tangent to the $y$-axis and returns to infinity tangent to the $x$-axis. Trajectories such as $y(x, 2)$ above $y(x, y^*)$ map onto loops inside the image of $y(x, y^*)$ and are 'pinched' at infinity—tangent to the $y$-axis at both ends. Trajectories such as $y(x, 1)$ below $y(x, y^*)$ also start out tangent to the $y$-axis but return tangent to the image of the negative $x$-axis." Christine Nowak (Austria) found it convenient, in fact, to start her solution by letting $x = 1/z$ and to then study the equation $dY/dz = (1/z)^2(1/Y - 1/z)$; in other words, to work near $x = \infty$ rather than near $x = 0$.

Also solved by Felix Albrecht, Preben Alsholm (Denmark), David Eberly, Kee-Wai Lau (Hong Kong), O. P. Lossers (The Netherlands), L. E. Mattics, Howard Morris, Chr. Nowak (Austria), R. I. Raichinov-Simeonov (Bulgaria), Lou Thurston, David V. V. Wend, and the proposer (jointly with J. Bethéry).

# REVIEWS

EDITED BY JOSEPH KONHAUSER
*Department of Mathematics, Macalester College, St. Paul, MN 55105*

*The Shape of Space.* By Jeffrey R. Weeks, Marcel Dekker, Inc., New York, 1985. x + 324 pp.

ALAN H. DURFEE
*Department of Mathematics, Mount Holyoke College, S. Hadley, MA 01075*

Many of the more spectacular mathematical advances of the twentieth century have occurred in the field of topology. Recently, much progress has been made in the area of three- and four-manifolds; other active areas are knot theory, algebraic topology, singularity theory, and dynamical systems, just to name a few. Fields medals have been given to Thurston, Donaldson, and Freedman. An interesting aspect of much recent work is that essential use is made of methods outside the field. Donaldson's work in four-manifolds, for example, uses results from mathematical physics; knot theory derives inspiration from representations of operator algebras; and three-manifold theory uses differential geometry.

Unfortunately, few undergraduate topology texts reflect recent activity. Most of these are on general topology and were, perhaps, written under the assumption that this foundational subject should be studied first. However, why recommend that an undergraduate mathematics major take general topology rather than, say, non-Euclidean or differential geometry, number theory, mathematical modeling or a host of other interesting courses? These subjects make an attractive end for an undergraduate education whereas general topology is the beginning of graduate work. Although general topology leads to the beautiful results of algebraic topology, the preparation is long and technical, and early applications are usually intuitively obvious and, hence, anticlimatic. Also, these results do not lead in the new directions mentioned above. Of course, the reliance of these new investigations on outside results also presents a special challenge to the author of an undergraduate text.

Even the most obscure topological proof usually has a simple geometric idea behind it, and a student who wants to prove such a theorem first needs to understand this geometry. However, visual intuition comes with experience, is usually not well developed in the typical undergraduate, and, hence, needs to be conveyed in the classroom. This can be tricky. A colleague of mine, for example, once gave as an examination question a complicated map of the surrounding area to be pasted together with various identifications. The result was supposed to be a two-holed torus, but he had made a mistake and instead a Klein bottle came out. Unfortunately, the Klein bottle had not yet been discussed in class.

Teaching topological intuition thus poses an interesting challenge. A good place to begin is with two-dimensional surfaces. *The Shape of Space*, for instance, starts as a sequel to *Flatland*: The scene is a two-dimensional universe in which lies a town called Flatsburgh. The inhabitants are plane figures, among them the protagonist, "A Square." Unlike the others who had not ventured far from town, A Square wanted to discover what lay beyond the horizon. One day he set off to the east.

encourage students to buy such a costly book, even though they would enjoy owning it. (The publishers apparently offer class discounts, though.)

Weeks' book on topology is highly recommended for a college mathematics course at any level.

*Native American Mathematics.* Edited by Michael P. Closs, University of Texas Press, Austin, 1986. 431 pp.

JAMES V. RAUFF
*Department of Mathematical and Computer Sciences, Millikin University, Decatur, IL 62522*

Standard texts and reference materials on the history of mathematics give only a perfunctory discussion of the development of mathematical thought in the pre-Columbian era ([**4**], for example). The Maya *bar and dot* notation is mentioned in most of these texts, and in some, a few remarks are made concerning Incan *quipus*. In general, however, the student of the history of mathematics is given a decidedly Eurocentric view of mathematical thinking, practice, and evolution. Recently, with the success of the decipherment of the Maya hieroglyphic writing [**5**] and advances in New World archeoastronomy [**3**] a serious examination of the mathematics of Native Americans has begun. *Native American Mathematics*, edited by the mathematician Michael P. Closs, is the first book to present the results of a wide range of this new research.

*Native American Mathematics* is a collection of papers ranging, mathematically, from counting to sophisticated calendrical computations, and ethnographically from the Inuit to the Inca. Although not all of the papers are completely successful in avoiding the ethnocentrism usually present in discussions of non-Western mathematics (see [**1**] for an excellent discussion of this problem), they are accessible to the mathematician and to the anthropologist and to students in both areas.

The 13 papers in *Native American Mathematics* may be grouped into seven categories:

(1) numeral systems as revealed by the spoken language (3 papers),
(2) counts recorded in rock art (1 paper),
(3) Mathematics and culture (3 papers),
(4) geometry and architecture (1 paper),
(5) Maya mathematics (2 papers),
(6) Aztec mathematics (2 papers),
(7) Inca mathematics (1 paper)

Here I will remark on only a sample of the 13 papers to give a general notion of the scope and feel of the collection.

J. Peter Denny, in his contribution "Cultural Ecology of Mathematics: Ojibway and Inuit Hunters," presents a detailed examination of counting, arithmetic, and geometry in these hunting societies from the point of view of the cultural context within which they are framed. Concerning counting, Denny argues that "counting is of less utility for hunters because most objects are known individually" (p. 137).

Marcia Ascher succinctly summarizes her work and that of Robert Ascher in "Mathematical Ideas of the Inca." The Aschers' [**2**] important analysis of the quipu is explained in a way that gets directly to the main insights of their work and also

tempts the reader towards further study. One unfortunate aspect of this paper is an editorial error. The rendering of the Guamon Poma de Ayala sketch of a purported Incan counting board given in Figure 10.1 of the Ascher paper is incorrect.

Michael Closs' survey of the intricacies of Maya calendrical calculations ("Mathematical Notation of the Maya") is perhaps the best short introduction to this fascinating area of study to date. Anyone beginning a study of Maya hieroglyphics and calendrics should start here.

On the other hand, A. Seidenberg's "The Zero in the Mayan Numerical Notation" is not so much a contribution to the history of mathematics as a polemic for diffusionism. The independent invention versus diffusion argument in anthropology has waxed and waned over the years and contemporary anthropologists generally consider the debate an oversimplification of rather complex processes. Regardless of the validity or advisability of Seidenberg's argument, this paper is out of place in this collection.

Finally, I would mention the short survey of the research in Mesoamerican geometry presented by Francine Vinette ("In Search of Mesoamerican Geometry"). Vinette clearly delineates the problems associated with uncovering ancient geometrical knowledge and outlines the progress made so far. The examples in this paper would be particularly useful for spicing up a class in plane mensuration geometry.

From Herbert Harvey and Barbara Williams' exciting discussion of Aztec area computation ("Decipherment and Some Implications of Aztec Numerical Glyphs") to Madison Beeler's report on "Chumash Numerals," *Native American Mathematics* takes the reader on a stimulating tour of the mathematics of the New World. Students of the history of mathematics, teachers of mathematics at all levels, and mathematicians interested in the interplay of mathematics and culture will find *Native American Mathematics* to be a treasure of examples, applications, interesting facts, and thought-provoking discussion. Every school and college library should have a copy of this splendid volume, and I am certain that more than a few teachers of mathematics will want their own personal copy.

REFERENCES

1. Marcia Ascher and Robert Ascher, Ethnomathematics, History of Science, 24 1986 125–144.
2. ______, The Code of the Quipu, University of Michigan Press, Ann Arbor, 1980.
3. Anthony Aveni, Native American Astronomy, University of Texas Press, Austin, 1977.
4. Howard Eves, An Introduction to the History of Mathematics, Holt, Rinehart and Winston, New York, 1969.
5. David Kelley, Deciphering the Maya Script, University of Texas Press, Austin, 1976.

*The Book of Prime Number Records*. By Paulo Ribenboim, Springer-Verlag, 1988. 476 + xxiii pp.

CARL POMERANCE
*Department of Mathematics, University of Georgia, Athens, GA 30602*

There is something about records of extremes that fascinates nearly everyone. On the evening weather broadcast, it is somehow comforting to find out that the miserable conditions outside are setting a new record. Think of how little would be left of a baseball telecast if statistics and talk of records were left out. In finance we

tempts the reader towards further study. One unfortunate aspect of this paper is an editorial error. The rendering of the Guamon Poma de Ayala sketch of a purported Incan counting board given in Figure 10.1 of the Ascher paper is incorrect.

Michael Closs' survey of the intricacies of Maya calendrical calculations ("Mathematical Notation of the Maya") is perhaps the best short introduction to this fascinating area of study to date. Anyone beginning a study of Maya hieroglyphics and calendrics should start here.

On the other hand, A. Seidenberg's "The Zero in the Mayan Numerical Notation" is not so much a contribution to the history of mathematics as a polemic for diffusionism. The independent invention versus diffusion argument in anthropology has waxed and waned over the years and contemporary anthropologists generally consider the debate an oversimplification of rather complex processes. Regardless of the validity or advisability of Seidenberg's argument, this paper is out of place in this collection.

Finally, I would mention the short survey of the research in Mesoamerican geometry presented by Francine Vinette ("In Search of Mesoamerican Geometry"). Vinette clearly delineates the problems associated with uncovering ancient geometrical knowledge and outlines the progress made so far. The examples in this paper would be particularly useful for spicing up a class in plane mensuration geometry.

From Herbert Harvey and Barbara Williams' exciting discussion of Aztec area computation ("Decipherment and Some Implications of Aztec Numerical Glyphs") to Madison Beeler's report on "Chumash Numerals," *Native American Mathematics* takes the reader on a stimulating tour of the mathematics of the New World. Students of the history of mathematics, teachers of mathematics at all levels, and mathematicians interested in the interplay of mathematics and culture will find *Native American Mathematics* to be a treasure of examples, applications, interesting facts, and thought-provoking discussion. Every school and college library should have a copy of this splendid volume, and I am certain that more than a few teachers of mathematics will want their own personal copy.

REFERENCES

1. Marcia Ascher and Robert Ascher, Ethnomathematics, History of Science, 24 1986 125–144.
2. ______, The Code of the Quipu, University of Michigan Press, Ann Arbor, 1980.
3. Anthony Aveni, Native American Astronomy, University of Texas Press, Austin, 1977.
4. Howard Eves, An Introduction to the History of Mathematics, Holt, Rinehart and Winston, New York, 1969.
5. David Kelley, Deciphering the Maya Script, University of Texas Press, Austin, 1976.

## *The Book of Prime Number Records*. By Paulo Ribenboim, Springer-Verlag, 1988. 476 + xxiii pp.

CARL POMERANCE
*Department of Mathematics, University of Georgia, Athens, GA 30602*

There is something about records of extremes that fascinates nearly everyone. On the evening weather broadcast, it is somehow comforting to find out that the miserable conditions outside are setting a new record. Think of how little would be left of a baseball telecast if statistics and talk of records were left out. In finance we

hear of the best and worst days on Wall Street, the richest people, the most expensive work of art, etc. In fact, one of the most successful books published (the Bible holds the record for the most successful) is the *Guinness Book of World Records*. This book even has a small section on mathematics, giving the largest number proved prime, the largest perfect number known, and a few other tidbits.

In fact, the *Guinness Book* just skims the surface. The entire subject of prime numbers is filled with records. For example, what is the largest known pair of twin primes? What is the least even number exceeding 2 not yet shown to be a sum of two primes? Or, what is the largest $k$ such that there exist $k$ primes forming an arithmetic progression? In his authoritative and charming book, Paulo Ribenboim answers these questions and many similar ones. We also find nine-and-a-half proofs that there are infinitely many primes, a discussion of possible functions representing primes, and problems, results and records concerning Mersenne primes, Fermat primes, regular primes, Sophie Germain primes, Wieferich primes, Wilson primes, and NSW primes (named for D. J. Newman, Daniel Shanks, and Hugh Williams).

Of course, not every mathematical statement is conveniently described in terms of a record. But the subject of prime numbers plays an unusual role within number theory and mathematics. There is a shadow side of the subject made up of heuristic arguments, giving prime number theory the flavor of an experimental science. With a few simple arguments from probability theory one is quickly led to heuristics for the twin prime conjecture, Goldbach's conjecture, the conjecture that there are only finitely many Fermat primes, the conjecture that there are infinitely many primes $p$ with $2^p \equiv 2 \pmod{p^2}$, and the conjecture that $\lim\sup(p_{n+1} - p_n)/\log^2 n = 1$ where $p_n$ denotes the $n$th prime, just to name a few. Yet every single one of these statements remains a very deep unsolved problem. Thus each may become the subject of a record, somehow measuring how close we have come to proving the conjecture.

For example, concerning the latter problem dealing with the maximal order of the gaps between consecutive primes, we do not even know that between consecutive squares there is always a prime. However, it is known that between all sufficiently large consecutive cubes there is a prime. What then is the record theorem? That is, what is the infimum of the set of $\theta$ for which it has been proved that $(*) p_{n+1} - p_n < p_n^\theta$ for all sufficiently large $n$ (depending on the choice of $\theta$)? Ribenboim reports that this record is held by C. J. Mozzochi who in 1986 showed $(*)$ for $\theta = 11/20 - 1/384 = .5473958\ldots$. But alas, glory is fleeting, and this record has already been broken by S.-t. Lou and Q. Yao, who have shown $(*)$ for all $\theta > 6/11 = .545454\ldots$. Again one is reminded of track records to the nearest hundredth of a second for 100 meters or the mile.

In the introduction Ribenboim writes that if he were to read in the newspaper of a brawl that had broken out in a local pub concerning the largest known pair of twin primes, he would find this highly civilized. He wryly suggests that his book might help settle such disputes more amicably. Of course, the publication of any book of records is a great stimulus to readers to try and break a record so as to get one's name in the book.

The subject of prime numbers is not merely a collection of interesting problems and results. It is a coherent theory. Ribenboim does an admirable job of synthesis in presenting a unified subject. Two central issues occupy more than half of the book. One of these is the subject of primality testing, i.e., how does one effectively

distinguish prime numbers from composite numbers? The other is the distribution of the prime numbers within the set of all natural numbers. Here the central result is the prime number theorem which shows that from a distant perspective, the prime numbers obey a very regular law. However, Ribenboim never lets us forget the apparent unpredictability and randomness of the primes when viewed close up.

The book is remarkably free of typographical errors and most of those remaining are innocuous. For example, on page 174 we see the set "$\{n \leq 1 | \varphi(n) \leq x\}$" where $\varphi$ is Euler's function. It is clear from the context that the first inequality is backwards. Of a more serious nature (there are really very few examples!), it is reported on page 79 that Bob Silverman had checked every prime $p \leq 150{,}000$ to see if $M_p = 2^p - 1$ is prime and after $p = 86{,}243$ had found only $p = 132{,}049$. In fact, Silverman did not do this search. In 1988, shortly after the book was published, W. N. Colquitt and L. Welsch, Jr. performed an exhaustive search up to the lower limit 132,049 and they did, in fact, discover a new Mersenne exponent! Namely, for $p = 110{,}503$, $M_p$ is prime.

The book does not contain proofs of major theorems nor even of most minor theorems. However, there are some proofs given "to break the rhythm," as Ribenboim puts it, and there are 100 pages of references. The level of the book is not very uniform. In some sections, class numbers of algebraic number fields are freely mentioned, while other sections are accessible to the proverbial bright high school student. This is not a book of exclusion; the reader is often prodded to skip around and find something else of interest if the current part gets too sticky. There are many levels on which this book may be enjoyed.

*The Book of Prime Number Records* takes now a firm place within the gallery of excellent elementary introductions to research in number theory including *Solved and Unsolved Problems in Number Theory* by D. Shanks, *Unsolved Problems in Number Theory* by R. K. Guy, and *Elementary Theory of Numbers* by W. Sierpinski. Ribenboim writes in his introduction that "Prime numbers are like cousins, members of the same family, resembling one another, but not quite alike." So too are these wonderful books. Thank you, Professor Ribenboim, for introducing us to a new cousin.

# TELEGRAPHIC REVIEWS

*Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.*

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

| | | |
|---|---|---|
| T: Textbook | P: Professional Reading | 1–4: Semesters |
| C: Computer Software | L: Undergraduate Library | **: Special Emphasis |
| S: Supplementary Reading | 13: Grade Level | ??: Questionable |

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

**Elementary, T(13: 1), S, L.** *Plane Trigonometry, Sixth Edition.* E. Richard Heineman. McGraw-Hill, 1988, xix + 297 pp, $35.95. [ISBN: 0-07-027935-7] A comprehensive, pre-calculus-level introduction to trigonometry, logarithms, and complex numbers, with a strong classical flavor. Combines time-tested and modern viewpoints—e.g., both interpolation in tables and use of calculators are treated. Emphasizes analytical, rather than computational, aspects: logarithms, e.g., are treated mainly as functions, not as computational devices. Many results, such as addition formulas, are proved using classical trigonometric techniques. Extensive appendices include answers to exercises, remarks on power series expansions, detailed tables of function values. (*Fourth Edition,* TR, December 1974.) PZ

**Mathematics Appreciation, S(16-18), P*, L*.** *From Cardinals to Chaos: Reflections on the Life and Legacy of Stanislaw Ulam.* Ed: Necia Grant Cooper. Cambridge U Pr, 1989, 319 pp, $75; $24.95 (P). [ISBN: 0-521-36494-9; 0-521-36734-4] A beautifully produced tribute to Stanislaw Ulam (1909-1984) and his wide-ranging contributions to mathematics, physics, and even biology. Buy two copies: one for the office, and one for the coffee table. BC

**Precalculus.** *Algebra and Trigonometry, Fourth Edition.* Dennis T. Christy. Wm C Brown, 1989, xix + 668 pp [ISBN: 0-697-05322-9]; *Fundamentals of Algebra and Trigonometry,* xvii + 621 pp [ISBN: 0-697-05323-7]; *College Algebra,* xvii + 476 pp. [ISBN: 0-697-05324-5] Continues the spiral approach of previous editions published as *Essentials of Precalculus Mathematics* (*Third Edition,* TR, February 1987). This edition includes a full chapter on systems of equations (including Gaussian elimination, matrix algebra, nonlinear systems, and linear programming), and a discrete algebra chapter (mathematical induction, counting techniques, and probability). *Fundamentals* permutes chapters and sections of the previous text to provide comprehensive coverage of algebra before trigonometry. *College Algebra* merely eliminates the two chapters on trigonometry from *Fundamentals.* JNC

**Precalculus, T(13: 1).** *Algebra and Trigonometry with Applications, Fourth Edition.* Bernard J. Rice, Jerry D. Strange. Brooks/Cole, 1989, xvi + 552 pp, $36. [ISBN: 0-534-10194-1] This edition retains the emphasis on functions, graphing, and the presentation of mathematical concepts in real-world situations. Changes include condensation of the review of basic algebra, new emphasis on calculators, and new sections on matrix algebra and polar coordinates. (*First Edition,* TR, January 1978.) JNC

**Precalculus, T(13: 1).** *Before Calculus: Functions, Graphs, and Analytic Geometry, Second Edition.* Louis Leithold. Harper & Row, 1989, xviii + 776 pp. [ISBN: 0-06-043911-4] This edition incorporates several revisions, e.g., chapter one has been condensed and complex numbers added, zeroes of a polynomial are now covered in one chapter, and the exercises have been revised and graded according to difficulty. (*First Edition,* TR, December 1985.) JNC

**Precalculus, T(13: 1).** *College Algebra and Trigonometry, Third Edition.* Robert Ellis, Denny Gulick. Harcourt Brace Jovanovich, 1989, vii + 632 pp, $29. [ISBN: 0-15-507932-8] Changes from earlier editions include the earlier introduction of linear equations and complex numbers, the elimination of discussion of the use of tables, a new section on compound and continuous interest, and some textual rewriting. (*First Edition,* TR, August-September 1981; *Second Edition,* TR, December 1984.) JS

**Precalculus, T*(13: 1).** *College Algebra: A Functions Approach, Fourth Edition.* Robert J. Mergener. Kendall/Hunt, 1989, viii + 372 pp, $28.95 (P). [ISBN: 0-8403-4968-8] Emphasizes functions (polynomial, rational, radical, logarithmic, and exponential) and their graphs. Lots of exercises, selected solutions in back. Begins with a pre-test followed by complete solutions. Workbook format includes trial problems, all solved in back. Many ex-

amples. Each chapter ends with applications section. Includes notes on calculator uses and abuses. Also includes sections on sequences and series, mathematical induction, conic sections, and elementary linear algebra. SB

**Precalculus, S.** *Mathematical Challenges: Puzzles and Problems in Secondary School Mathematics.* Mathematical Education on Merseyside. Janson, 1989, ix + 59 pp, $17.50 (P). A collection of problems that have been used in high school competitions in northwest England. JNC

**Precalculus, T(13).** *Plane Trigonometry, Third Edition.* R. David Gustafson, Peter D. Frisk. Brooks/Cole, 1989, xv + 442 pp, $35.25. [ISBN: 0-534-09822-3] Changes from previous edition (TR, December 1982) include expanded function discussion, more emphasis on general solutions to trigonometric equations, an analytic geometry chapter, more applications. Covers standard trigonometry material as well as appendices covering geometry, exponential and logarithmic functions, and linear interpolation. Written at a fairly elementary level. Introduces trigonometric functions as "degree" functions, *then* introduces radian measure. Many exercises; selected solutions in back. SB

**Finite Mathematics, T(13).** *Finite Mathematics with Applications.* David E. Zitarelli, Raymond F. Couglin. Saunders College, 1989, xx + 563 pp, $33. [ISBN: 0-03-011292-3] A text covering bits of linear algebra, linear programming, set theory, and probability and statistics. Abounds in applications of these subjects to modern, everyday life. ES

**Finite Mathematics, T.** *Mathematics in Business Administration.* Yves Nievergelt. Irwin, 1989, xvii + 494 pp, $34. [ISBN: 0-256-06914-X] Designed for executives entering an MBA program. Covers mathematical skills required and used in graduate quantitative business courses: geometry and algebra review, functions (with emphasis on graphs, slopes, and areas), algebraic geometry of linear and quadratic equations, mathematics of finance, differential calculus for optimization. Includes actual case-study applications from *The Wall Street Journal.* End of each section includes self-test with solutions and answers to all exercises. SB

**Education, T*, P*, L.** *Secondary Mathematics Instruction: An Integrated Approach.* Margaret A. Farrell, Walter A. Farmer. Janson, 1988, xix + 360 pp, $42.50 (P). [ISBN: 0-939765-26-8] Revised 1980 edition of *Systematic Instruction in Mathematics for the Middle and High School Years.* Designed for both in-service and pre-service teachers. Interactive format. Integrates theory, practice, and research. New edition includes references to research which supports specific suggestions or results. Each chapter begins with an introductory "advanced organizer" section with college students in mind, followed by an introductory activity (mostly laboratory or interview tasks, but some paper-and-pencil activities), ends with summary and set of simulation or practice activities, and suggested further readings. Chapters include modes of instruction, feedback, adolescent reasoning patterns, structure and learning of mathematics, design and evaluation of instructional strategies, resources, and disciplines. SB

**Education, P*, L.** *Fantastiks of Mathematiks: Applications of Secondary Mathematics.* Cliff Sloyer. Janson, 1986, xi + 143 pp, $13.95 (P). [ISBN: 0-939765-00-4] Forty problems and discussion guides designed to motivate students and teachers to work on a variety of mathematical applications. Before each problem is stated, the mathematics background required is listed. Although designed for teachers of secondary mathematics, these notes could also be useful for teachers of "non-technical" college mathematics. These nonstandard, interesting applications could be very helpful in generating student discussion on problem solving strategies. Subjects range through algebra, inequalities (e.g., arithmetic-geometric mean), systems of equations, graphing, combinations, probability, geometry, trigonometry, sequences and series, calculus, and elementary linear programming. SB

**Education, L.** *Nonroutine Problems: Doing Mathematics.* Robert London. Janson, 1989, iv + 60 pp, $18.50 (P). [ISBN: 0-939765-30-6] Ten problems concerning area calculation, pi, infinity, functions, and optimization designed to develop problem solving strategies for students with two years of college preparatory mathematics. Includes pedagogical notes to help teachers structure the assignments, generate class discussion, and evaluate results. Each problem also includes suggestions for further investigation and examples of student solutions. The problems are intended to require at least a few hours work over a two-week period. SB

**Education, T(13-17), S, L.** *The Complete Problem Solver, Second Edition.* John R. Hayes. Lawrence Erlbaum Associates, 1989, xvii + 357 pp, $59.95; $24.95 (P). [ISBN: 0-89859-782-X; 0-8058-0309-2] Not everyone who is able to solve a problem is equally adept at explaining their thinking. This book is concerned with thinking about thinking. It is designed to help readers improve their general problem solving abilities, and to provide information about the psychology of problem solving (e.g., how does the mind work?). Divided into four parts: Problem Solving (representation, planning, evaluation); Memory and Knowledge Acquisition; Decision Making; and Creativity and Invention. Especially valuable for teachers. (*First Edition,* TR, February 1983.) LCL

**Education, S, P.** *LOGO: Theory & Practice.* Dennis Harper. Brooks/Cole, 1989, xviii + 409 pp, $33.75 (P). [ISBN: 0-534-09720-0] Presents Logo to students and in-service teachers and summarizes the views of researchers, teachers, educational leaders, and children; includes explanations, hints, projects, and activities. JNC

**History, S(16), L*.** *Unrolling Time: Christiaan Huygens and the Mathematization of Nature.* Joella G. Yoder. Cambridge U Pr, 1988, xi + 238 pp,

$42.50. [ISBN: 0-521-34140-X] Case study of the inter-relationship between mathematics and physics in the work of Christiaan Huygens (1629-1695): on Huygens' research on the pendulum clock, the constant of gravitational acceleration, centrifugal force, and the theory of evolutes: his relationship with other scientists; priority disputes: the *Horologium Oscillatorium;* his role in the rise of applied mathematics: his work pattern and dependence upon outside stimuli; his consistent appeal to and his trust in mathematics. Adheres to original derivations and avoids the finished propositions. Original drawings supplemented by redrawn versions for sake of clarity. Modern notation introduced in some geometric arguments. Copious notes. Numerous references. JK

**Foundations, T*(14-15: 1).** *Introduction to Mathematical Structures.* Steven Galovich. Harcourt Brace Jovanovich, 1989, xii + 484 pp, $27. [ISBN: 0-15-543468-3] What every student should know before taking upper-level courses in mathematics and computer science. Goal is to provide an introduction to mathematical structure and mathematical thinking. Minimal prerequisite is one-variable calculus. Familiarity with multivariable calculus and linear algebra is helpful but not absolutely necessary. Clear and readable. Readers are encouraged to engage "yourself with the book." Loaded with insightful remarks. Chapter-end comments, historical remarks, and references add perspective and provide direction for further study. Excellent exercise and problem sets—routine though challenging. JK

**Graph Theory, P.** *Graph Theory in Memory of G.A. Dirac.* Ed: Lars Døvling Andersen, *et al.* Annals of Disc. Math., V. 41. North-Holland (US Distr: Elsevier Science), 1989, xiv + 517 pp, $122. [ISBN: 0-444-87129-2] Forty-two papers from participants (and others) of a meeting at Sandbjerg, Denmark, June 1985, in memory of the graph theorist G.A. Dirac. Includes a list of sixteen problems posed at a problem session. LC

**Discrete Mathematics, T(14: 1).** *Discrete Mathematics with Applications.* William Barnier, Jean B. Chan. West, 1989, xiv + 488 pp, $44.50. [ISBN: 0-314-45966-9] Covers sets, algorithms and logic, directed graphs and relations, relations and functions, combinatorics and finite probability, logic and proof, integers and binary numbers, Boolean algebra, graphs and trees, recurrence relations, and automata theory and formal languages. Each chapter includes an application and optional programming notes and exercises (using Pascal). JNC

**Discrete Mathematics, T*(13: 1).** *Introduction to Discrete Mathematics.* Robert J. McEliece, Robert B. Ash, Carol Ash. Random House, 1989, xv + 514 pp, $37.50. [ISBN: 0-394-35819-8] Uses a unifying theme of algorithms to present chapters on set theory, combinatorics, graphs, propositional calculus, Boolean algebra, models for computing machines, formal languages, mathematical games, discrete probability, and finite difference equations. Includes chapter overviews and summaries, and historical footnotes. Test bank available. JNC

**Discrete Mathematics, T(13-14: 1).** *An Introduction to Discrete Mathematics, Second Edition.* Steven Roman. Harcourt Brace Jovanovich, 1989, x + 469 pp, $33. [ISBN: 0-15-541730-4] Additions to the *First Edition* (TR, October 1986; Extended Review, June-July 1987) include the strong form of induction, second order linear recurrence relations, and elementary probability theory. LC

**Number Theory, P*.** *Number Theory, Trace Formulas and Discrete Groups.* Ed: Karl Egil Aubert, Enrico Bombieri, Dorian Goldfeld. Academic Pr, 1989, xx + 510 pp, $69.95. [ISBN: 0-12-067570-6] Eight survey articles and twenty-one research summaries from a symposium held July 14-21, 1987 in Oslo. Honoring Atle Selberg's seventieth birthday, the papers focus on automorphic forms and trace formulae, and on prime numbers and the Riemann zeta function. They demonstrate Selberg's profound influence in these areas. GG

**Number Theory, T**(14-16: 1).** *Elementary Number Theory, Second Edition.* David M. Burton. Wm C Brown, 1989, xvii + 450 pp. [ISBN: 0-697-05919-7] Second edition of a very thorough, readable text which introduces many ideas in their historical context. Includes historical sketches on early number theory and some of its major contributors. No formal prerequisites; accessible to students with or without knowledge of abstract algebra. Main changes from the 1980 *Revised Printing* (TR, December 1980; *First Edition,* TR, March 1976): new sections on cryptography, Merten's conjecture, primes in arithmetic progression, absolute pseudoprimes, amicable numbers; expanded Fermat primes section including recent numerical updates; 150 new exercises. SB

**Linear Algebra, T*(14-15: 1, 2), L.** *Linear Algebra, Second Edition.* Stephen H. Friedberg, Arnold J. Insel, Lawrence E. Spence. Prentice-Hall, 1989, xiv + 530 pp. [ISBN: 0-13-537102-3] A refinement of the 1979 *First Edition* (TR, October 1979) with lots of minor but careful changes in the exposition. A book which emphasizes structure rather than computation. Chapter titles show this: vector spaces, linear transformations and matrices, elementary matrix operations and systems of linear equations, determinants, diagonalization, inner-product spaces, and canonical forms. JAS

**Linear Algebra, T(13-14: 1), L.** *Elementary Linear Algebra.* William L. Perry. McGraw-Hill, 1988, xiii + 495 pp, $35.95. [ISBN: 0-07-049431-2] In terms of topics chosen and organization, the book is very much in the spirit of the new linear algebra texts aimed at the sophomore level. Distinctive features include earlier and more extensive discussion of numerical methods and some attention to the geometrical properties of linear transformations. Complex numbers are used. JS

**Algebra, P.** *Deformation Theory of Algebras and Structures and Applications.* Ed: Michiel Hazewinkel, Murrav Gerstenhaber. NATO ASI Ser.

C, V. 247. Kluwer Academic, 1988, viii + 1030 pp, $199. [ISBN: 90-277-2804-6] Based on the advanced study institute held in 1986 at Il Ciocco, Castelvecchio-Pascoli, Italy. Contains nineteen papers, one nearly a book, on deformations for many kinds of algebras over general fields, as well as special cases of Lie algebras, and the algebras of functional analysis. JAS

**Algebra, T(18: 1, 2), P.** *Field Theory: Classical Foundations and Multiplicative Groups.* Gregory Karpilovsky. Pure & Appl. Math., V. 120. Marcel Dekker, 1988, viii + 551 pp, $99.75. [ISBN: 0-8247-8029-9] Extensive summary of classical results in field theory along with many relatively recent refinements and extensions concerning multiplicative groups of fields. Assumes one-year graduate algebra course. SB

**Algebra, T(18: 1, 2), P.** *Topics in Field Theory.* Gregory Karpilovsky. Math. Stud., V. 155. North-Holland (US Distr: Elsevier Science), 1989, xi + 546 pp, $144.75. [ISBN: 0-444-87297-3] Follow-up to first-year graduate algebra course. Covers separable algebraic extensions, transcendental extensions, derivations, purely inseparable extensions, Galois theory, Abelian extensions and radical extensions, including co-galois theory. Small overlap with author's *Field Theory: Classical Foundations and Multiplicative Groups* (see TR above), but has different emphasis. SB

**Algebra, P.** *Lecture Notes in Mathematics-1348: Categorical Algebra and its Applications.* Ed: F. Borceux. Springer-Verlag, 1988, viii + 375 pp, $32.90 (P). [ISBN: 0-387-50362-5] A selection of twenty-eight papers presented at the first Louvain-la-Neuve Conference on Categorical Algebra and Its Applications, held in the summer of 1988. List of participants. JS

**Algebra, T(15-16), L.** *A First Course in Abstract Algebra, Fourth Edition.* John B. Fraleigh. Addison-Wesley, 1989, ix + 518 pp, $40.80. [ISBN: 0-201-16847-2] Fairly extensive revision from earlier editions. Includes a more traditional organization into eight longer chapters, some use of linear algebra and matrices in examples, functions written to the left of the argument, and addition of some brief historical notes. (*First Edition,* TR, August-September 1969; *Third Edition,* TR, January 1983.) JS

**Algebra, T(18), P.** *Lectures on the Asymptotic Theory of Ideals.* D. Rees. London Math. Soc. Lect. Note Ser., V. 113. Cambridge U Pr, 1988, ix + 202 pp, $24.95 (P). [ISBN: 0-521-31127-6] An exposition of developments through the early-to-mid 1980's. Important topics include Matijevic's theorem, various valuation theorems, and the theory of mixed multiplicities. SG

**Calculus, T?(13-14: 1-3).** *Basic Technical Mathematics with Calculus, Second Edition.* Peter Kuhfittig. Brooks/Cole, 1989, xxii + 1076 pp, $43.25. [ISBN: 0-534-10062-7] Ambitious text attempting to cover algebra, linear algebra, trigonometry, analytic geometry, single variable calculus, infinite series, differential equations, and Laplace transforms. Nice use of two-color graphics, but many sections seem short and rushed. Fairly routine exercises. Emphasizes techniques and applications over concepts and rigor. (*First Edition,* TR, January 1985.) SB

**Calculus, T(13).** *Calculus with Applications.* Claudia Dunham Taylor, Lawrence Gilligan. Brooks Cole, 1989, xiv + 700 pp, $40.50. [ISBN: 0-534-10272-7] Calculus for students majoring in biology, business, economics, and the social sciences. Most of the applications are business-oriented. Aside from basic calculus the text includes chapters on differential equations and probability. SG

**Calculus, T(13).** *Calculus with Analytic Geometry: Late Trigonometry Version, Third Edition.* Howard Anton. Wiley, 1989, xviii + 1393 pp, $57.68. [ISBN: 0-471-62211-7] No mention of trigonometric functions until Chapter 8. Identical to standard *Third Edition* except for placement of trigonometry material. Problems in first seven chapters modified accordingly. (*First Edition,* TR, April 1981; *Second Edition,* TR, November 1984; Extended Review, March 1986; *Third Brief Edition,* TR, February 1989.) SB

**Calculus, T(13: 3).** *Calculus with Analytic Geometry.* Leonard I. Holder. Wadsworth, 1988, xviii + 964 pp. [ISBN: 0-534-08202-5] A pleasantly concise and readable presentation of traditional topics which introduces concepts intuitively and relegates proofs to end of sections and chapters; avoids an extended review of precalculus, and incorporates trigonometric functions early. Includes a section on probability density functions and chapters on vector field theory and differential equations. JNC

**Calculus, T(13: 1, 2), S.** *Guide to Analysis.* F. Mary Hart. Macmillan Education, 1988, xi + 202 pp, £8.95 (P). [ISBN: 0-333-43788-8] Written to accompany a first-year course at a British university. Starts with the study of sequences and series, then turns to limits, continuity, and derivatives. Closes with topics related to differentiation such as maxima and minima, power series. Does not cover integration. GN

**Calculus, T(13: 1).** *Calculus with Applications.* Raymond F. Coughlin, David E. Zitarelli. Saunders College, 1989, xvii + 601 pp, $34. [ISBN: 0-03-011283-4] An applications-oriented calculus text with an impressive number of in-depth case studies, applied examples and exercises, especially from business/economics, but also biology, ecology, and the social sciences. These are cataloged in a separate index for easy reference. Chapters 1-7 constitute a basic core in the differential-integral calculus of polynomial and exponential functions, with latter chapters on multivariable calculus, differential equations, trigonometric functions, Taylor series, and probability via calculus. Referenced case studies include demographic effects on the social security system, and reliability analysis in the space shuttle disaster. CE

**Real Analysis, T(14).** *After Calculus: Analysis.* David J. Foulis, Mustafa A. Munem. Dellen, 1989,

xvii + 556 pp, $35. [ISBN: 0-02-339130-8] Intended as a text for the transitional course from "cookbook" freshman calculus to "more rigorous and sophisticated" upper-level mathematics courses. The text covers logic, basic set theory; relations; algebraic, order, and topological properties of the reals, differentiation, integration, and series. Historical notes and annotated bibliography included after each chapter. SG

**Complex Analysis, S(18), P.** *Lecture Notes in Mathematics-1345: Holomorphic Dynamics.* Ed: X. Gomez-Mont, J. Seade, A. Verjovski. Springer-Verlag, 1988, 321 pp, $28.60 (P). [ISBN: 0-387-50226-2] Seventeen research papers—a few in French—on complex dynamical systems, from the Second International Colloquium on Dynamical Systems held in Mexico in July 1986. One long paper, by Branner and Douady, describes the technique of "surgery" on parameter spaces which arise from viewing holomorphic polynomials as dynamical systems. PZ

**Complex Analysis, S(18), P.** *Several Complex Variables III: Geometric Function Theory.* Ed: G.M. Khenkin. Ency. of Math. Sci., V. 9. Springer-Verlag, 1989, 261 pp, $59. [ISBN: 0-387-17005-7] Seven long research-expository papers, by Soviet mathematicians, summarizing history and recent developments in various research areas of several complex variables: entire functions, value distribution theory, invariant metrics, holomorphic mappings, complex geometry. With extensive bibliographies. PZ

**Complex Analysis, S(18), P.** *Lecture Notes in Mathematics-1351: Complex Analysis, Joensuu 1987.* Ed: I. Laine, S. Rickman, T. Sorvali. Springer-Verlag, 1988, xv + 378 pp, $32.90 (P). [ISBN: 0-387-50370-6] Thirty-five research papers on complex analysis in one variable; partial proceedings of the XIIIth Rolf Nevanlinna Colloquium held in Finland in 1987. The volume honors, and many papers reflect, the work of Lars V. Ahlfors. Topics include extremal functions, quasiconformal mappings, Teichmüller spaces, and complex dynamics. PZ

**Differential Equations, T(14-15: 1).** *Elementary Differential Equations, Seventh Edition.* Earl D. Rainville, Phillip E. Bedient. Macmillan, 1989, xiii + 546 pp, $34. [ISBN: 0-02-397860-0] An old friend back for another round. Still without frills. Its popularity rests in large part on its direct style. Covers all the basic methods of solving ordinary differential equations. Plenty of examples and mostly routine exercises (with answers to almost all that ask for one). Additions include a derivation of Kepler's laws from those of Newton. Flexible. Also, chapter 1-16 (through nonlinear equations but not power series) appear separately in a brief version. (*Fourth Edition,* TR, June-July 1969; *Fifth Edition,* TR, December 1974; *Sixth Edition,* TR, August-September 1981.) JK

**Differential Equations, S(18), P.** *Exponential Type Calculus for Linear Delay Equations.* S.M. Verduyn Lunel. CWI Tract, V. 57. Mathematisch Centrum, 1989, 125 pp, Dfl. 19 (P). [ISBN: 90-6196-364-8] Central theme is the asymptotic behavior of the solutions of linear delay equations, and, more generally, linear autonomous (translation invariant) systems with some aspect of age taken into account. Includes a "short course" on Riemann-Stieltjes integral and methods of the Laplace transformation. Covers exponential type calculus for a class of entire functions useful in the derivation of convergence criteria for Fourier type expansions. A few examples at the end; over 40 references, several to recent works. JK

**Partial Differential Equations, P.** *Navier-Stokes Equations.* Peter Constantin, Ciprian Foias. Lect. in Math. Ser. U of Chicago Pr, 1988, ix + 190 pp, $14.95 (P); $34.95. [ISBN: 0-226-11549-6; 0-226-11548-8] Studies classical topics such as existence, regularity and uniqueness results, vanishing viscosity limits, analyticity, backward uniqueness, and eigenvalue asymptotes. GN

**Numerical Analysis, P.** *Domain Decomposition Methods.* Ed: Tony F. Chan, *et al.* SIAM, 1989, 463 pp, $49.50 (P). [ISBN: 0-89871-233-5] Contains the proceedings of the Second International Symposium on Domain Decomposition Methods held at UCLA in January 1988. The book is divided into four parts: theory, algorithms, parallel implementation, and applications. SM

**Numerical Analysis, T(16), L.** *The Numerical Solution of Ordinary and Partial Differential Equations.* Granville Sewell. Academic Pr, 1988, xii + 271 pp, $39.95. [ISBN: 0-12-637475-9] Covers some of the more common finite difference methods for solving both ordinary and partial differential equations. The final chapter introduces some of the basics of the finite element method. Assumes some knowledge of numerical analysis but not necessarily a background in ordinary or partial differential equations. Examples include programs written in Fortran. GN

**Operator Theory, P.** *Introduction to the Spectral Theory of Polynomial Operator Pencils.* A.S. Markus. Transl. of Math. Mono., V. 71. AMS, 1988, iv + 250 pp, $95. [ISBN: 0-8218-4523-3] A translation from the 1986 Russian edition, this monograph centers on work of Keldysh on multiple completeness of the eigenvectors of a pencil of operators and asymptotic behavior of its eigenvalues. An introductory chapter is followed by chapters on Keldysh pencils, factorization, and self-adjoint pencils. Comments on the literature, bibliography. JS

**Functional Analysis, P.** *Modular Function Spaces.* Wojciech M. Kozlowski. Pure & Appl. Math., V. 122. Marcel Dekker, 1988, viii + 252 pp, $89.75. [ISBN: 0-8247-8001-9] An introduction to the theory of modular function spaces, i.e., function spaces defined by certain linear functionals which are not norms. Extends work by Orlicz, Nakano, Luxemburg, and others. Applications to theory of summation of integrals; finding maximal domains of continuity for nonlinear operators; and analytic extension of functions of several complex variables. ES

**Functional Analysis, T(17-18: 1), P.** *Lec-*

*ture Notes in Mathematics-1364: Convex Functions, Monotone Operators and Differentiability.* Robert R. Phelps. Springer-Verlag, 1989, ix + 115 pp, $14.30 (P). [ISBN: 0-387-50735-3] An up-to-date account of results on the differentiability of convex functions on infinite dimensional spaces, related in ways explained in the Preface to J. Giles, *Convex Analysis with Application to Differentiation of Convex Functions,* Pitman, 1982, and R. Bourgin, *Geometric Aspects of Convex Sets with the Radon-Nikodym Property,* Springer-Verlag, 1983. AWR

**Functional Analysis.** *Analyse Fonctionnelle: Avec Exercices.* Hervé Lehning. Math. Supérieures et Spéciales, V. 5. Masson, 1988, 248 pp, (P). [ISBN: 2-225-81226-8] A text intended for students in the first three years of study in a French university and thus roughly corresponds to a first year graduate text in the United States. Main themes are function spaces, functions defined by series to integrals, and differential equations. Many interesting, extended exercises. SG

**Functional Analysis, T(16-17: 1).** *Elements of Functional Analysis, Second Edition.* I.J. Maddox. Cambridge U Pr, 1988, xii + 242 pp, $59.50; $22.95 (P). [ISBN: 0-521-35350-5; 0-521-35868-X] Understandable introduction to functional analysis with lots of interesting examples and exercises. Covers basic topology and analysis plus standard results concerning maps between normed linear spaces, duality, Banach algebras, and Hilbert space. Ends with a chapter on applications to differential equations and summability (*First Edition,* TR, May 1970; Extended Review, March 1971). ES

**Functional Analysis, P.** *Lecture Notes in Mathematics-1344: Potential Theory, Surveys and Problems.* Ed: J. Král, *et al.* Springer-Verlag, 1988, viii + 270 pp, $24.30 (P). [ISBN: 0-387-50210-6] Proceedings of a conference in Prague, July 1987. Concerns potential theory and its applications to the theory of functions, differential equations, and probability theory. Contains a list of eighteen open problems in the field with commentary by their proposors. ES

**Functional Analysis, P.** *Lecture Notes in Mathematics-1363: Tsirelson's Space.* Peter G. Casazza, Thaddeus J. Shura. Springer-Verlag, 1989, viii + 204 pp, $21.10 (P). [ISBN: 0-387-50678-0] Tsirelson's Space is a reflexive Banach space with monotone unconditional Schauder basis and no embedded copies of $c_0$ or $l_p$. These notes collect what is known about this space. LC

**Functional Analysis, P.** *Banach Space Theory.* Ed: Bor-Luh Lin. Contemp. Math., V. 85. AMS, 1989, xiv + 521 pp, $48 (P). [ISBN: 0-8218-5092-x] Proceedings from a workshop on Banach Space Theory held at the University of Iowa, July 5-25, 1987. LC

**Functional Analysis, S(18), P.** *Lectures on Integral Transforms.* N.I. Akhiezer. Transl. of Math. Mono., V. 70. AMS, 1988, v + 108 pp, $45. [ISBN: 0-8218-4524-1] A series of lecture notes concerning various types of integral transforms (especially the Fourier) on spaces of integrable and square-integrable functions of several variables. Applications to the theory of harmonic and analytic functions, the moment problem, orthogonal polynomials, and problems in mathematical physics. ES

**Functional Analysis, P.** *Lecture Notes in Mathematics-1356: Multiparameter, Eigenvalue Problems and Expansion Theorems.* Hans Volkmer. Springer-Verlag, 1988, vi + 157 pp, $16.30 (P). [ISBN: 0-387-50479-6] Summary of classical results in multiparameter eigenvalue theory for Hermitian matrices, compact Hermitian operators, and semibounded, self-adjoint operators. Accessible to readers familiar with one-parameter eigenvalue problems for compact Hermitian operators. Extends several theorems by weakening "definiteness" requirements. Supplements Atkinson's book on same subject. ES

**Analysis, P.** *Lecture Notes in Mathematics-1365: Topics in Calculus of Variations.* Ed: M. Giaquinta. Springer-Verlag, 1989, x + 196 pp, $21.10 (P). [ISBN: 0-387-50727-2] Survey lectures on maps with singularities (H. Brezis), free boundary problems (L.A. Caffarelli), minimal foliations on a torus (J. Moser), nonlinear problems (L. Nirenberg), total scalar curvature (R.M. Schoen), and Teichmüller theory (A.J. Tromba). BC

**Algebraic Geometry, T(18: 1, 2), P*.** *Lecture Notes in Mathematics-1358: The Red Book of Varieties and Schemes.* David Mumford. Springer-Verlag, 1988, v + 309 pp, $28.60 (P). [ISBN: 0-387-50497-4] Reprint of notes distributed in mid-1960's by Harvard mathematics department. Quick, informal introduction to basics of algebraic geometry. Very readable reference, especially for non-algebraic geometers to learn the basic definitions in the language of schemes. Assumes understanding of Galois theory, separability, transcendence, ring localization, ideal decomposition in Noetherian rings and integral dependence. SB

**Differential Geometry, P.** *Theory of Multicodimensional $(n + 1)$-Webs.* Vladislav V. Goldberg. Math. & Its Applic. Kluwer Academic, 1988, xxi + 466 pp, $139. [ISBN: 90-277-2756-2] The first text on the subject written in English. The author presents the basic theory and many applications. The results presented in the book were for the most part originally proved by the author. SG

**Algebraic Topology, P.** *Algebraic Homotopy.* Hans Joachim Baues. Cambridge U Pr, 1988, xix + 466 pp, $89.50. [ISBN: 0-521-33376-8] A presentation of an axiomatic framework for homotopical algebra via the concept of a cofibration category. JAS

**Algebraic Topology, T(17), P.** *Topological Fields.* Witold Więsław. Pure & Appl. Math., V. 119. Marcel Dekker, 1988, vii + 309 pp, $99.75. [ISBN: 0-8247-7731-X] "Presents most of the known results dealing with topological fields." Begins with fundamentals and later discusses Krull valuations, topologies of type V, locally compact fields, connected fields, connections of arithmetic and topology, and extensions of topologies. Extensive bibliography. BH

**Algebraic Topology, P.** *The Homology of Hopf Spaces.* Richard M. Kane. Math. Lib., V. 40. North-Holland (US Distr: Elsevier Science), 1988, xv + 479 pp, $105.25. [ISBN: 0-444-70464-7] An exposition of the theory of finite $H$-spaces which has developed over the last thirty years. This book brings together many topics previously found only in the research literature. RB

**Differential Topology, P.** *Lecture Notes in Mathematics-1366: Grassmannians and Gauss Maps in Piecewise-linear Topology.* Norman Levitt. Springer-Verlag, 1989, v + 203 pp, $21.10 (P). [ISBN: 0-387-50756-6] An extension to the PL case of a number of bundle-oriented tools of differential topology. JAS

**Topology, P.** *Partition Problems in Topology.* Stevo Todorcevic. Contemp. Math., V. 84. AMS, 1989, xi + 116 pp, $22 (P). [ISBN: 0-8218-5091-1] Deals with a topological version of the Ramsey problem for the uncountable. Simple cases of this problem include, for example, the famous Souslin problem and the well-known problem from topological measure theory asking if all regular Radon measures are $\sigma$-finite. JAS

**Topology, C, P, L.** *Computers in Geometry and Topology.* Ed: Martin C. Tangora. Lect. Notes in Pure & Appl. Math., V. 144. Marcel Dekker, 1989, viii + 317 pp, $99.75 (P). [ISBN: 0-8247-8031-0] A collection of papers ranging from fun graphics to studies of computational complexity and REDUCE algorithms for computing the cohomology of nilpotent groups. JAS

**Topology, T*(15-16: 1), L.** *Principles of Topology.* Fred H. Croom. Saunders College, 1989, xi + 312 pp. [ISBN: 0-03-012813-7] Very readable text; nice use of graphics; many examples. Begins with an intuitive, historical introduction along with various prerequisite material. Then covers topological concepts beginning with $R^1$ and $R^2$, then metric spaces, emphasizing Euclidean space and Hilbert space. Covers standard topology syllabus: basis, subbasis, topological equivalence, topological invariants, separability, connectedness, compactness, compactification, product and quotient spaces, manifolds, separation axioms, and metrizability. Final chapter introduces algebraic topology and includes an introduction to groups in the appendix. Each chapter ends with historical notes of interest and suggested further reading. Each section begins with an introductory paragraph for motivation and ends with a number of exercises. SB

**Topology, T(16-17: 2).** *Topology: A Geometric Account of General Topology, Homotopy Types and the Fundamental Groupoid.* Ronald Brown. Math. & its Applic. Halsted Pr, 1988, xviii + 460 pp, $59.95. [ISBN: 0-470-21217-9] A thorough revision of a 1968 McGraw-Hill text *Elements of Modern Topology* (TR, January 1969; Éxtended Review, May 1969). Begins, as before, with about 100 pages of standard point set topics followed by almost another hundred pages on cell complexes and appropriate topological tools and examples. The distinguishing feature of this book is the development of the fundamental groupoid. It is this last part that has been most revised to reflect suggested improvements in the exposition and deeper insights resulting from more recent research. JAS

**Operations Research, S(17), P.** *Logic-Based Decision Support: Mixed Integer Model Formulation.* Robert G. Jeroslow. Annals of Disc. Math., V. 40. North-Holland (US Distr: Elsevier Science), 1989, xv + 222 pp, $78. [ISBN: 0-444-87119-5] Based on ten lectures given by the author at the Advanced Research Institute on Discrete Applied Mathematics at Rutgers University. Discusses the connections between mixed integer programming problems, decision support systems, and intelligent systems. SM

**Operations Research, T(16), S, P, L.** *Cooperative Games, Solutions and Applications.* Theo Driessen. Theory & Decision Library, Ser. C. Kluwer Academic, 1988, xiv + 222 pp, $74. [ISBN: 90-277-2729-5] Focuses on the theory of cooperative games in characteristic function form. The theoretical aspects of both modeling and solution are detailed. The $\tau$-value and $k$-convex games are discussed. Introductory first chapter, with examples from various disciplines. No exercises. SM

**Operations Research, T(15-17: 1), L.** *Introduction to the Mathematics of Operations Research.* Kevin J. Hastings. Pure & Appl. Math., V. 128. Marcel Dekker, 1989, viii + 407 pp, $99.75. [ISBN: 0-8247-8039-6] Develops the mathematics of operations research before presenting the applications. Therefore, the emphasis is on the mathematics rather than on presenting a string of techniques tailored to specific applications. The mathematical rigor here exceeds that found in typical undergraduate operations research texts. Plenty of exercises, no solutions given. Expensive. SM

**Operations Research, T(15: 2), L.** *Linear Programming.* James Calvert, William Voxman. Harcourt Brace Jovanovich, 1989, xi + 655 pp, $34. [ISBN: 0-15-551027-4] Usual basic topics: examples, simplex method, tableau, revised simplex, duality, sensitivity, integer programming, and transportation models. Emphasizes examples and how to formulate LP's rather than abstraction and rigor. Newer methods mentioned but not shown. Software available. TH

**Optimization, P.** *Theory of Suboptimal Decisions: Decomposition and Aggregation.* A.A. Pervozvanskii, V.G. Gaitsgori. Math. & Its Applic. Kluwer Academic, 1988, xvii + 384 pp, $99. [ISBN: 90-277-2401-6] Develops mathematically rigorous procedures for the composition and aggregation of mathematical programs where the resulting decompositions or aggregations are not fully equivalent to the original program. This leads naturally to the use of perturbation methods for the approximate solution of deterministic mathematical programs, stochastic programs, and optimal control problems. SM

**Control Theory, T(17: 1), S, P, L.** *An Introduc-*

*tion to Hankel Operators.* Jonathan R. Partington. Math. Soc. Stud. Texts, V. 13. Cambridge U Pr, 1988, 103 pp, $12.95 (P). [ISBN: 0-521-36791-3] A readable treatment of Hankel operators, which crop up in control theory as well as functional analysis. Includes a lovely quote from John Dickson Carr: "I am a mathematician, sir. I never permit myself to think." BC

**Control Theory, S(18), P.** *Lecture Notes in Control and Information Sciences-118: Singular Control Systems.* L. Dai. Springer-Verlag, 1989, ix + 332 pp, $44 (P). [ISBN: 0-387-50724-8] Sums up the development of singular control system theory. Linear algebra and linear systems theory are assumed. Prerequisites are briefly covered in the appendices. SM

**Systems Theory, P, L.** *Lecture Notes in Control and Information Sciences-116: Abstract Systems Theory.* M.D. Mesarovic, Y. Takahara. Springer-Verlag, 1989, viii + 439 pp, $56 (P). [ISBN: 0-387-50529-6] Abstract systems theory seeks to identify and formalize concepts common to a vast range of systems—biological, political, mechanical, etc. It can be thought of as a "system theory of systems theories." Reasonably self-contained. Includes new results. SM

**Probability, P.** *Lecture Notes in Statistics-51: Extreme Value Theory.* Ed: J. Hüsler, R.-D. Reiss. Springer-Verlag, 1989, x + 279 pp, $29.70 (P). [ISBN: 0-387-96954-3] Twenty-two papers from December 1987 conference in Oberwolfach, West Germany. Extremes of special stochastic processes, rate of convergence and finite expansions, probabilistic aspects concerning records, extremes of dependent random variables, role of point processes, inference about domain of attraction, models for univariate and multivariate extremes. TH

**Probability, P.** *Lecture Notes in Mathematics-1362: École d'Été de Probabilités de Saint-Flour XV-XVII, 1985-87.* P. Diaconis, *et al.* Springer-Verlag, 1988, 459 pp, $45.10 (P). [ISBN: 0-387-50549-0] Six series of lectures from the years 1985-1987. JAS

**Probability, S(17-18), P, L.** *Gibbs Measures and Phase Transitions.* Hans-Otto Georgii. Stud. in Math., V. 9. Walter de Gruyter, 1988, xiv + 525 pp, DM 178. [ISBN: 3-11-010455-5] A clearly written, readable treatment of Gibbs measure, which describes the equilibrium states of physical systems with a large number of interacting components. Motivated by physics (phase transitions correspond to non-uniqueness of the measure), the theory leads to interesting questions in probability. BC

**Stochastic Processes, P.** *Probability Approximations via the Poisson Clumping Heuristic.* David Aldous. Appl. Math. Sci., V. 77. Springer-Verlag, 1989, xv + 269 pp, $44. [ISBN: 0-387-96899-7] "Problems about random extrema can often be translated into problems about sparse random sets which resemble i.i.d. random clumps." This heuristic is exploited to discuss a large number and variety of problems in theoretical and applied probability. A lot of ideas and information packed tightly together. Contains an extensive bibliography. TAV

**Stochastic Processes, S(18).** *Statistical Inference from Stochastic Processes.* Ed: N.U. Prabhu. Contemp. Math., V. 80. AMS, 1988, xv + 386 pp, $38 (P). [ISBN: 0-8218-5087-3] A collection of seventeen papers from invited speakers at the Joint Summer Research Conference on Statistical Inference from Stochastic Processes held at Cornell University, summer 1987. Papers presented covered five broad topics: foundations, counting processes and survival analysis, likelihood, applications to statistics and probability models, and processes in economics. JJ

**Statistics, P?** *Mathematical Statistics with Application in Flood Hydrology.* J. Reimann. Akademiai Kiado, 1989, 330 pp, $34. [ISBN: 963-05-4832-1] Intermediate level mathematical statistics: distributions, estimation, hypothesis testing, confidence intervals, Markov chains, regression. Examples from a Hungarian river system. Missing hydrology topics such as extreme value distributions and simulation. TH

**Computer Literacy, T.** *Inside Data Processing: Computers and their Effective Use in Business.* Alastair de Watteville. Chartwell-Bratt, 1988, ix + 173 pp, (P). [ISBN: 0-86238-181-9] An introductory text for business students in a technical institute. A strictly expository treatment of such topics as "What is a computer?" and "What is a programming language?" (examples cited: COBOL, Fortran, and RPG's, as well as BASIC and Pascal). JAS

**Programming, S.** *Advanced Programming with Microsoft QuickC.* Keith Weiskamp. Academic Pr, 1989, xiii + 549 pp, $49.95. [ISBN: 0-12-742684-1] Intermediate programming with lots of examples for a compiler that supports the ANSI standard in a PC world. JAS

**Programming, C*, P*, L*.** *C Traps and Pitfalls.* Andrew Koenig. Addison-Wesley, 1989, xi + 147 pp, $16.25 (P). [ISBN: 0-201-17928-8] A lovely collection of enticing puzzles (with explanatory solutions) which really explain the inner workings of C. If you can laugh at yourself over a silly mistake which kept you up until three a.m., you will see the humor in this delightfully written little book. If you program in C, you need this book—even if you won't see the humor. JAS

**Programming, C, P, L.** *C Programming in a UNIX Environment.* Judy Kay, Bob Kummerfeld. Intern. Comput. Sci. Ser. Addison-Wesley, 1989, xii + 340 pp, (P). [ISBN: 0-201-12912-4] A systematic introduction—with exercises—to C and its UNIX Version 7 (System V, 4BSD) environment including libraries, the preprocessor, make, and lint. Treats ANSI standards in an appendix. JAS

**Programming, T(13-14: 1, 2), C.** *ADA from the Beginning.* Jan Skansholm. Intern. Comput. Sci. Ser. Addison-Wesley, 1988, xiii + 617 pp, $29.25 (P). [ISBN: 0-201-17522-3] An extensive introduction to programming with ADA assuming no previous programming experience. Also suitable as

a "soft" introduction for people with some programming experience. Does not cover the parallel programming aspects of ADA. Appears to be a leisurely, clear exposition. JAS

**Algorithms, T*(16: 1), S, L*.** *Algorithmics: Theory and Practice.* Gilles Brassard, Paul Bratley. Prentice-Hall, 1988, xvi + 361 pp. [ISBN: 0-13-023243-2] A readable, yet detailed, examination of the algorithms of classical computer science, including searching, sorting, trees and graphs, matrix multiplication, and number theory. Later chapters on preconditioning, probabilistic algorithms, and complexity. Sections begin with general description of particular algorithms, then interlace examples of applications with exercises. GG

**Computer Science, P.** *Translating Programs into Delay-Insensitive Circuits.* J.C. Ebergen. CWI Tract, V. 56. Mathematisch Centrum, 1989, x + 216 pp, Dfl. 33 (P). [ISBN: 90-6196-363-X] It is shown that the design of delay-insensitive circuits can be reduced to the design of programs. A syntax-directed translation of programs into delay-insensitive connections of basic elements. Includes many examples to illustrate this technique. CEC

**Computer Science, T*(14-16: 1, 2), L.** *Data Structures.* Rick Decker. Prentice-Hall, 1989, xiv + 417 pp. [ISBN: 0-13-198813-1] A nicely written introductory text using Pascal. Emphasizes abstract data types and presents the material in a way suitable for a second course in computer science, immediately following any of a number of common introductory courses. JAS

**Applications, T(15-16: 1), S, L*.** *Mathematical Models and Their Analysis.* Frederic Y.M. Wan. Harper & Row, 1989, xvi + 394 pp. [ISBN: 0-06-046902-1] Following a one-chapter introduction, each of remaining fourteen chapters is devoted to one of ten independent problems, ranging over many non-stochastic techniques. Prerequisites are calculus, a touch of linear algebra, ordinary differential equations, and, for a few chapters, partial differential equations. First model for a problem is often simple, with its flaws used to develop more sophisticated models. Chapters conclude with related computational exercises. GG

**Applications (Biological Science), L*.** *The Geometry of Genetics.* A.M. Findley, S.P. McGlynn, G.L. Findley. Mono. in Chemical Physics. Wiley, 1989, xi + 156 pp, $44.95. [ISBN: 0-471-05617-0] An interdisciplinary search for "hidden symmetries" in molecular genetics. Mappings, groups, and morphisms are tools for study of structure—the genetic code; linear and metric spaces for statics—the replication of DNA; and differential geometry for dynamics—the evolution of genes. All necessary mathematics and biology is introduced as needed. LAS

**Applications (Economics), P.** *Lecture Notes in Economics and Mathematical Systems-322: The Rational Expectations Equilibrium Inventory Model: Theory and Applications.* Ed: Tryphon Kollintzas. Springer-Verlag, 1989, xi + 269 pp, $30.30 (P). [ISBN: 0-387-96940-3] Six papers at the research level on aggregate inventory behavior. Four are from a June 1987 meeting held at Wesleyan University. GG

**Applications (Engineering), P.** *Generation and Application of Pseudorandom Sequences for Random Testing.* V.N. Yarmolik, S.N. Demidenko. Wiley, 1988, vii + 167 pp, $49.95. [ISBN: 0-471-91999-3] Translation from Russian. Generation of pseudorandom numbers for testing of complex engineering systems, particularly digital electrical apparatus. Emphasis is on hardware solutions and testing of physical apparatus, not computer modeling. Circuit diagrams are provided. TH

**Applications (Engineering), P.** *Systems with Hysteresis.* M.A. Krasnosel'skiĭ, A.V. Pokrovskiĭ. Trans: Marek Niezgódka. Springer-Verlag, 1989, xviii + 410 pp, $79. [ISBN: 0-387-15543-0] Develops new general methods for describing and analyzing systems with hysteresis nonlinearities. The approach is based on the concept of the hysteron, a deterministic transducer of input (e.g., deformation) into output (e.g., stress), as the elementary carrier of hysteresis effects in more complex systems. Careful attention is given to the physical origin of the models studied. CE

**Applications (Physics), S(18).** *Lecture Notes in Statistics-48: Bayesian Spectrum Analysis and Parameter Estimation.* G. Larry Bretthorst. Springer-Verlag, 1988, xii + 209 pp, $23.40 (P). [ISBN: 0-387-96871-7] An extensive revision of author's Ph.D. dissertation concerning the parameter estimation model in spectral analysis (frequencies, chirp rates, decay rates, signals). Uses Bayesian estimation techniques to include prior information on estimates. Discusses designing experiments, collecting data, and analyzing data. Several applications for physicists, economists, and engineers. Background required: mathematics needed in graduate physics, probability, and statistics. JJ

### Reviewers

RJA: Richard J. Allen, St. Olaf; DFA: David F. Appleyard, Carleton; SB: Steve Benson, St. Olaf; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; CE: Christopher Ennis, Carleton; SG: Steven Galovich, Carleton; GG: George Gilbert, St. Olaf; BH: Bruce Hanson, St. Olaf; RH: Rhonda Hatcher, St. Olaf; TH: Timothy Hesterberg, St. Olaf; JJ: Jason Jones, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; GN: Gail Nelson, Carleton; AO: Arnold Ostebee, St. Olaf; SP: Samuel Patterson, Carleton; MR: Matthew Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MS: Myriam Steinback, Macalester; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; LW: Lynne Walling, St. Olaf; PZ: Paul Zorn, St. Olaf.

# THREE IMPORTANT BOOKS ON MATHEMATICS EDUCATION

## RESPONSES TO THE CHALLENGE: KEYS TO IMPROVED INSTRUCTION BY TEACHING ASSISTANTS AND PART-TIME INSTRUCTORS

The Committee on Teaching Assistants
and Part-Time Instructors,
Bettye Anne Case, Chair

The committee that prepared this volume has been gathering information on policies, practices, successes, failures, and goals connected with the use of teaching assistants and part-time instructors. In this volume the committee presents and analyzes data showing who these teachers are, the extent and nature of their teaching duties, and the efforts made to assimilate them into the faculties. This volume will help you to see how your department compares nationally, to decide what steps you and your school should take, and to understand what additional resources might be needed.

280 pp., 1988, ISBN-0-88385-061-3
List: $15.00
Catalog Number NTE-11

## GUIDELINES FOR THE CONTINUING MATHEMATICAL EDUCATION OF TEACHERS

Committee on the Mathematical Education of Teachers

These guidelines will be very useful to school teachers and supervisors, to college administrators who plan continuing edcucation programs, to the college teachers who design and teach courses for teachers, and to school administrators who must think about requirements for continuing education of teachers. The guidelines are rich in specifics on course content, giving clear objectives for all courses. Teachers who want to dig out material for themselves or in order to enrich their classes will find the more than 500 references provided here under various topics an invaluable aid.

90 pp., 1988, ISBN-0-88385-060-5
List: $8.00
Catalog Number NTE-10

## THE USE OF CALCULATORS IN THE STANDARDIZED TESTING OF MATHEMATICS

John W. Kenelly, Editor

The calculator is a universal tool for all those involved in quantitative work from science and engineering to business. Routine use of calculators is part of the training and testing of students headed for these fields. But this is not yet the case in mathematics. This symposium, jointly sponsored by the MAA and The College Board, sets out clearly the theoretical and practical issues that must be addressed as calculators are brought more fully into the mathematics curriculum. This is a practical group concerned with specific tests and test items. General theoretical considerations are set off by the specifics of individual test items and students' success rates on them. The Ohio Early College Mathematics Placement Test is reported on in detail by Joan R. Leitzel and Bert K. Waits. James W. Wilson and Jeremy Kilpatrick examine the theoretical issues in the development of calculator-based tests. John Harvey, now Chair of the MAA's Committee on Placement Examinations looks at the issues surrounding calculator use on placement examinations, as well as giving an overview of the symposium and a survey of developments through 1988.

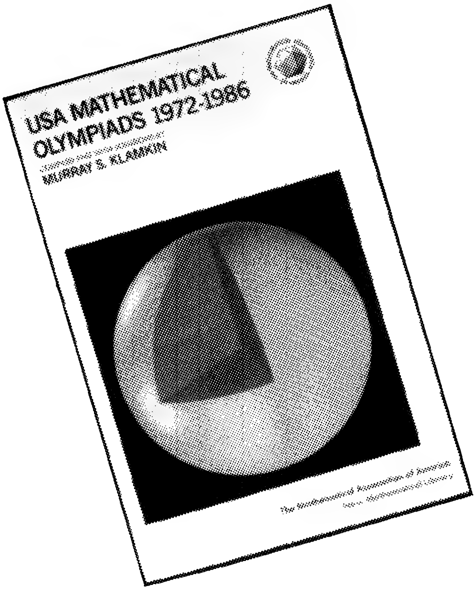
vi + 50 pp., 1989, Copublished by the MAA and The College Board. LC No. 88-064-100

List: $6.50 Member: $6.50

*Order from:*

The Mathematical Association of America
1529 Eighteenth Street, N. W.
Washington, D. C. 20036
(202) 387-5200

# USA MATHEMATICAL OLYMPIADS




Every year 100 of the most mathematically talented high school students in the country compete in the USA Mathematical Olympiad (USAMO). The USAMO is the third stage of a three-tiered mathematical competition for high school students in the United States and Canada that begins with the AHSME taken by over 400,000 students, continues with the American Invitational Mathematics Exam involving 2,000 students, and culminates in the 100-contestant USAMO.

## USA MATHEMATICAL OLYMPIADS 1972-1986, PROBLEMS AND SOLUTIONS

Compiled by Murray S. Klamkin

People delight in working on problems "because they are there, " for the sheer pleasure of meeting a challenge. This is a book full of such delights. In it, Murray S. Klamkin brings together 75 original USA Mathematical Olympiad (USAMO) problems for years 1972-1986, with many improvements, extensions, finger exercises, open problems, references and solutions, often showing alternative approaches. The problems are coded by subject and solutions are arranged by subject as an aid to those interested in a particular field. Contains a glossary of frequently used terms and theorems, and a comprehensive bibliography with items numbered and referred to in brackets in the text. The problems are intriguing and the solutions elegant and informative. Students and teachers will enjoy working these challenging problems. Indeed all those those who are mathematically inclined will find many delights and pleasant challenges in this book.

180 pp., l988, ISBN-0-88385-634-4

List: $13.50 MAA Member: $12.50

Catalog Number NML-33

*Order from:*

The Mathematical Association of America
1529 Eighteenth Street, N. W.
Washington, D. C. 20036
(202) 387-5200

# SIAM BOOKSHELF

## Stochastic Processes in the Neurosciences
*Henry C. Tuckwell*

This monograph is centered on quantitative analysis of nerve-cell behavior. The work is foundational, with many higher order problems still remaining, especially in connection with neural networks. Thoroughly addressed topics include stochastic problems in neurobiology, and the treatment of the theory of related Markov processes.

**Contents.** Deterministic theories and stochastic phenomena in neurobiology; synaptic transmission; early stochastic models for neuronal activity; discontinuous Markov processes with exponential decay; one-dimensional diffusion processes; stochastic pde's; statistical analysis of stochastic neural activity; channel noise; Wiener kernel expansions; stochastic activity of neuronal populations.

126 pages, softcover ISBN 0-89871-232-7
CBMS/NSF Regional Conference Series in Applied Mathematics 1989
List Price $24.50 **SIAM/CBMS Member Price: $19.60**
**Order Code CB56**

## Combinatorial Algorithms: An Update
*Herbert S. Wilf*

This volume focuses on some of the exciting, vigorous new work being done in this field, including recent developments in algorithms for generating combinatorial objects such as partitions, Gray codes, and trees. It is a comprehensive update of Wilf and Albert Nijenhuis' book, *Combinatorial Algorithms*, which was first published in 1975 and revised in 1978. This new monograph is based on a series of ten lectures given by Wilf at the CBMS-NSF Conference on Selection Algorithms for Combinatorial Objects held in 1987 at Colorado College.

**Contents.** Preface; The Original Gray Code; Other Gray Codes; Variations on the Theme; Choosing 2-Samples; Listing Rooted Trees; Random Selection of Trees; Listing Free Trees; Generating Random Graphs; Bibliography.

56 pages, softcover, ISBN 0-89871-231-9
CBMS-NSF Regional Conference Series in Applied Mathematics 1989
List Price $12.00 **SIAM/CBMS Member Price 9.60**
**Order Code CB55**

## Mathematics Applied to Deterministic Problems in the Natural Sciences
*C.C. Lin and L.A. Segel*

"...The words 'applied mathematics' decorate the covers of many books these days, but precious little of the practice of applied mathematics seeps through into the pages bound between those covers. Of the select few that do indeed introduce the reader to applied mathematics, this book may well be the best. I know of none better."
— Paul Davis, *American Mathematical Monthly*, 1975 review of the original printing.

This reprint includes multipart exercises and an annotated bibliography, corrects all known misprints, and updates the original printing with remarks and references.

**Contents.** Overview -- What is Applied Mathematics? Deterministic Systems and Ordinary Differential Equations; Random Processes and Partial Differential Equations; Superposition, Heat Flow, and Fourier Analysis; Further Developments in Fourier Analysis; Some Fundamental Procedures -- Simplification Dimensional Analysis and Sealing; Regular Perturbation Theory; Illustration of Techniques on a Physiological Flow Problem; Introduction to Singular Perturbation Theory; Singular Perturbation Theory Applied to a Problem in Biochemical Kinetics; Three Techniques Applied to the Simple Pendulum; Continuous Fields -- Longitudinal Motion of a Bar; The Continuous Medium; Field Equations of Continuum Mechanics; Inviscid Fluid Flow; Potential Theory.

609 pages, softcover, ISBN 0-89871-229-7
Classics in Applied Mathematics Series 1988
List Price: $24.00 **SIAM Member Price: $19.20**
**Order Code CL01**

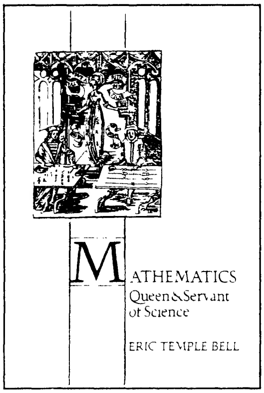
## An Algorithmic Theory of Numbers, Graphs and Convexity
*Laszlo Lovasz*

This book is a study of how complexity questions in computing interact with classical mathematics in the numerical analysis of issues in algorithm design. Two algorithms are studied in detail -- the ellipsoid method and the simultaneous diophantine approximation method. Algorithmic designers concerned with linear and nonlinear combinatorial optimization will find this volume especially useful.

91 pages, softcover, ISBN 0-89871-203-3
CBMS-NSF Regional Conference Series in Applied Mathematics 1986
List Price $14.00 **SIAM/CBMS Member Price $11.20**
**Order Code CB50**

**To Order: Call toll-free 1-800-447-SIAM, or send payment and a copy of this ad (with price and order code of book(s) you wish to purchase circled) to SIAM, Dept. BKAM89, P.O. Box 7260, Philadelphia, PA 19101-7260.** CBMS discount applies to members of: ALMATC, AMS, ASA, ASL, ASSM, AWM, IMS, MAA, NCSM, NCTM, ORSA, SIAM, SOA, and TIMS. Please circle the group(s) of which you are a member.

# MATHEMATICS & BIOGRAPHY




## MATHEMATICS: QUEEN AND SERVANT OF SCIENCE

E.T. Bell

An absorbing account of pure and applied mathematics from the geometry of Euclid to that of Riemann and its application in Einstein's theory of relativity. The twenty chapters treat such topics as: algebra, number theory, logic, probability, infinite sets and the foundations of mathematics, rings, matrices, transformations, groups, geometry, and topology. Republished in 1987 with corrections and an added Foreward by Martin Gardner.

454 pp., ISBN 0-88385-446-3
Paperbound

List: $15.95 MAA Member: $11.95

Catalog Number QAS

*The book deserves a place in today's market. It is a much more popular work than most histories of the subject, and that is exactly what makes it accessible to students as well as to nonmathematicians. It is rewarding reading for . . . teachers and students at all mathematical levels.*

Morris Kline of The Courant Institute

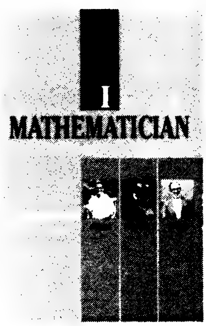
## I WANT TO BE A MATHEMATICIAN
## An Automathography in Three Parts

Paul R. Halmos

This is a book to be read with interest by all those who know, or might want to know, what mathematicians and mathematical careers are like. Paul R. Halmos begins with his school days and carries the reader swiftly through a career that has sustained itself at a high level since his first post-doctoral days at the Institute for Advanced Study in 1939, where he worked with John von Neumann among others. Still going strong in 1988, Halmos has contributed much to logic, operator theory, ergodic theory, and the literature in general.

442 pp., 1988, Paper, ISBN 0-88385-445-7

List: $18.00 MAA Member: $15.00




Catalog Number IWM

*It is a truly unique book, which nobody but Paul Halmos could have written. I think it will be a classic.*

Constance Reid

*The book is exciting, witty, and well worth the time invested in its study. It communicates what it means to be a mathematician.*

John Dossey in *The Mathematics Teacher*

*Order from:*

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036
(202) 387-5200

THE AMERICAN
# MATHEMATICAL MONTHLY

Volume 96, Number 8 October 1989

## Contents

(ISSN 0002-9890)

ARTICLES

Pi, Euler Numbers, and Asymptotic Expansions . . . . . . . . . . . J. M. BORWEIN, P. B BORWEIN, AND K. DILCHER 681

The William Lowell Putnam Mathematical Competition LEONARD F. KLOSINSKI, G. L. ALEXANDERSON, AND LOREN C. LARSON 688

Geometry of Continued Fractions . . . . . . . . . . M. C. IRWIN 696

EDITOR'S CORNER

The White Screen Problem . . . . . . . . HERBERT S. WILF 704

LETTERS TO THE EDITOR . . . . . . . . . . . . . . . . . . . . . 708

NOTES

Fifty Years of Putnam Trivia . . . . . . . . . . . . JOSEPH A. GALLIAN 711

Fixed Points of the Twisted Cyclic Shift Operator . . . . . LARRY W. CUSICK AND PETER TANNENBAUM 713

Fubini's Theorem for Null Sets . . . . . ERIC K. VAN DOUWEN 718

An Optimization Problem . . . . . . . . . RICHARD BASSEIN 721

A Characterization of a Class of Composition Operators . . . . . . . . . . . . R. E. LEWKOWICZ 725

THE TEACHING OF MATHEMATICS

Coupled Linear Differential Equations with Real Coefficients . . . . . MOGENS E LARSEN AND BJARNE S. JENSEN 727

Irrationals and the Fundamental Theorem of Arithmetic . . . . . . . . . . . . . . DAVID J. SPROWS 732

A Supplement to I. N. Herstein's Remark on Finite Fields . . . . . . . . . . . . . HARAGAURI N. GUPTA 733

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions . . . . . . . . . . . . . . . . 734

Advanced Problems and Solutions . . . . . . . . . . . . . . . . 743

REVIEWS

The Pólya Picture Album—Encounters of a Mathematician edited by Gerald L. Alexanderson . . . . . . . . . . . . . . FRANK HARARY 750

Geometric Measure Theory. A Beginner's Guide by Frank Morgan . . . . . . . . . . . . . . FREDERICK J. ALMGREN, JR. 753

TELEGRAPHIC REVIEWS . . . . . . . . . . . . . . . . . . . . . 757

## NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to ANITA E. SOLOW, Department of Mathematics, Grinnell College, Grinnell, IA 50112; in geometry or topology to DENNIS DETURCK, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Three copies of proposed problems or solutions should be sent to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, University of Calgary, Alberta, Canada T2N 1N4.




Reprint Permission: MARCIA P. SWARD, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: MS. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $26 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $29. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $90 per year.





Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

# Pi, Euler Numbers, and Asymptotic Expansions

J. M. Borwein, P. B. Borwein, and K. Dilcher[1], *Dalhousie University, Halifax, Canada*

Jonathan M. Borwein was an Ontario Rhodes Scholar (1971) at Jesus College, Oxford, where he completed a D. Phil. (1974) with Michael Dempster. Since 1974 he has worked at Dalhousie University where he is professor of mathematics. He has also been on faculty at Carnegie-Mellon University (1980–82). He was the 1987 Coxeter-James lecturer of the Canadian Mathematical Society and was awarded the Atlantic Provinces Council on the Sciences 1988 Gold Medal for Research. His research interests include functional analysis, classical analysis, and optimization theory.

Peter B. Borwein obtained a Ph.D. (1979) from the University of British Columbia, under the supervision of David Boyd. He spent 1979–80 as a NATO research fellow in Oxford. Since then he has been on faculty at Dalhousie (except for a sabbatical year at the University of Toronto) and is now Associate Professor of Mathematics. His research interests include approximation theory, classical analysis, and complexity theory.

Karl Dilcher received his undergraduate education and a Dipl. Math. degree at Technische Universität Clausthal in West Germany. He completed his Ph.D (1983) at Queen's University in Kingston, Ontario with Paulo Ribenboim. Since 1984 he has been teaching at Dalhousie, where he is now Assistant Professor. His research interests include Bernoulli numbers and polynomials, and classical complex analysis.

**1. Introduction.** Gregory's series for $\pi$, truncated at 500,000 terms, gives to forty places

$$4\sum_{k=1}^{500,000}\frac{(-1)^{k-1}}{2k-1} = 3.14159\underline{0}6535897932\underline{40}4626433832\underline{6}9502884197.$$

The number on the right is not $\pi$ to forty places. As one would expect, the 6th digit after the decimal point is wrong. The surprise is that the next 10 digits are correct. In fact, only the 4 underlined digits aren't correct. This intriguing observation was sent to us by R. D. North [10] of Colorado Springs with a request for an explanation. The point of this article is to provide that explanation. Two related

[1] Research of the authors supported in part by NSERC of Canada.

examples, to fifty digits, are

$$\begin{aligned}\frac{\pi}{2} &\doteq 2\sum_{k=1}^{50{,}000}\frac{(-1)^{k-1}}{2k-1}\\ &= 1.5707\underset{1}{\underline{8}}632679489\underset{-1}{\underline{7}}619231321\underset{5}{\underline{1}}9163975\underset{-61}{\underline{205}}209858\underline{33147388}\end{aligned}$$

and

$$\begin{aligned}\log 2 &\doteq \sum_{k=1}^{50{,}000}\frac{(-1)^{k+1}}{k}\\ &= .6931\underset{1}{\underline{3}}7180\underset{-1}{\underline{6}}59945309\underset{2}{\underline{39}}72321214\underset{-16}{\underline{74}}1765680\underset{272}{\underline{483}}00134\underline{46572},\end{aligned}$$

where all but the underlined digits are correct. The numbers under the underlined digits are the numbers that must be added to correct these. The numbers 1, $-1$, 5, $-61$ are the first four *Euler numbers* while 1, $-1$, 2, $-16$, 272 are the first five *tangent numbers*. Our process of discovery consisted of generating these sequences and then identifying them with the aid of Sloane's *Handbook of Integer Sequences* **[11]**. What one is observing, in each case, is an asymptotic expansion of the error in Euler summation. The amusing detail is that the coefficients of the expansion are integers. All of this is explained by Theorem 1.

The standard facts we need about the *Euler numbers* $\{E_i\}$, the *tangent numbers* $\{T_i\}$, and the *Bernoulli numbers* $\{B_i\}$, may all be found in **[1]** or in **[6]**. The numbers are defined as the coefficients of the power series

$$\sec z = \sum_{n=0}^{\infty}(-1)^n\frac{E_{2n}z^{2n}}{(2n)!}, \tag{1.1}$$

$$\tan z = \sum_{n=0}^{\infty}(-1)^{n+1}\frac{T_{2n+1}z^{2n+1}}{(2n+1)!} \qquad \text{and } T_0 = 1, \tag{1.2}$$

$$\frac{z}{e^z-1} = \sum_{n=0}^{\infty}\frac{B_n z^n}{n!}. \tag{1.3}$$

They satisfy the relations

$$\sum_{k=0}^{n}\binom{2n}{2k}E_{2k} = 0, \qquad E_{2n+1} = 0, \tag{1.4}$$

$$B_n = \frac{-nT_{n-1}}{2^n(2^n-1)} \qquad n \geqslant 1, \tag{1.5}$$

and

$$\sum_{k=0}^{n}\binom{n+1}{k}B_k = 0. \tag{1.6}$$

These three identities allow for the easy generation of $\{E_n\}$, $\{T_n\}$, and $\{B_n\}$. The

first few values are recorded below.

| $n$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| $E_n$ | 1 | 0 | $-1$ | 0 | 5 | 0 | $-61$ | 0 | 1365 |
| $T_n$ | 1 | $-1$ | 0 | 2 | 0 | $-16$ | 0 | 272 | 0 |
| $B_n$ | 1 | $\frac{-1}{2}$ | $\frac{1}{6}$ | 0 | $\frac{-1}{30}$ | 0 | $\frac{1}{42}$ | 0 | $\frac{-1}{30}$ |

It is clear from (1.4) that the Euler numbers are integral. From (1.5) and (1.6) it follows that the tangent numbers are integers. Also,

$$|E_{2n}| \sim \frac{4^{n+1}(2n)!}{\pi^{2n+1}} \quad \text{and} \quad |B_{2n}| \sim \frac{2(2n)!}{(2\pi)^{2n}}$$

as follows from (5.1) and (5.2) below. The main content of this note is the following theorem. The simple proof we offer relies on the Boole Summation Formula, which is a pretty but less well-known analogue of Euler summation. The details are contained in Sections 2 and 3 (except for c] which is a straightforward application of Euler summation). More complicated developments can be based directly on Euler summation or on results in [**9**].

THEOREM 1. *The following asymptotic expansions hold:*

a] $$\frac{\pi}{2} - 2\sum_{k=1}^{N/2} \frac{(-1)^{k-1}}{2k-1} \sim \sum_{m=0}^{\infty} \frac{E_{2m}}{N^{2m+1}} = \frac{1}{N} - \frac{1}{N^3} + \frac{5}{N^5} - \frac{61}{N^7} + \cdots$$

b] $$\log 2 - \sum_{k=1}^{N/2} \frac{(-1)^{k-1}}{k} \sim \frac{1}{N} + \sum_{m=1}^{\infty} \frac{T_{2m-1}}{N^{2m}} = \frac{1}{N} - \frac{1}{N^2} + \frac{2}{N^4} - \frac{16}{N^6} + \frac{272}{N^8} - \cdots$$

and

c] $$\frac{\pi^2}{6} - \sum_{k=1}^{N-1} \frac{1}{k^2} \sim \frac{1}{2N^2} + \sum_{m=0}^{\infty} \frac{B_{2m}}{N^{2m+1}} = \frac{1}{N} + \frac{1}{2N^2} + \frac{1}{6N^3} - \frac{1}{30N^5} + \frac{1}{42N^7} \cdots .$$

From the asymptotics of $\{E_n\}$ and $\{B_n\}$ and (1.5) we see that each of the above infinite series is everywhere divergent; the correct interpretation of their asymptotics is

a'] $$\sum_{m=1}^{\infty} \frac{E_{2m}}{N^{2m}} = \sum_{m=1}^{K} \frac{E_{2m}}{N^{2m}} + 0\left(\frac{(2K+1)!}{(\pi N)^{2K+1}}\right)$$

b'] $$\sum_{m=1}^{\infty} \frac{T_{2m-1}}{N^{2m}} = \sum_{m=1}^{K} \frac{T_{2m-1}}{N^{2m}} + 0\left(\frac{(2K+1)!}{(\pi N)^{2K-1}}\right)$$

c'] $$\sum_{m=1}^{\infty} \frac{B_{2m}}{N^{2m+1}} = \sum_{m=1}^{K} \frac{B_{2m}}{N^{2m+1}} + 0\left(\frac{(2K+1)!}{(2\pi N)^{2K+1}}\right),$$

where in each case the constant concealed by the order symbol is independent of $N$ and $K$. In fact, the constant 10 works in all cases.

**2. The Boole Summation Formula.** The Euler polynomials $E_n(x)$ can be defined by the generating function

$$\frac{2e^{tx}}{e^t + 1} = \sum_{n=0}^{\infty} E_n(x) \frac{t^n}{n!} \qquad (|t| < \pi); \tag{2.1}$$

(see [**1**, p. 804]). Each $E_n(x)$ is a polynomial of degree $n$ with leading coefficient 1. We also define the periodic Euler function $\bar{E}_n(x)$ by

$$\bar{E}_n(x+1) = -\bar{E}_n(x)$$

for all $x$, and

$$\bar{E}_n(x) = E_n(x) \quad \text{for} \quad 0 \leqslant x < 1.$$

It can be shown that $\bar{E}_n(x)$ has continuous derivatives up to the $(n-1)$st order.

The following is known as **Boole's summation formula** (see, for example, [**9**, p. 34]).

LEMMA 1. *Let $f(t)$ be a function with $m$ continuous derivatives, defined on the interval $x \leqslant t \leqslant x + \omega$. Then for $0 \leqslant h \leqslant 1$*

$$f(x + h\omega) = \sum_{k=0}^{m-1} \frac{\omega^k}{k!} E_k(h) \cdot \frac{1}{2}\left(f^{(k)}(x+\omega) + f^{(k)}(x)\right) + R_m,$$

*where*

$$R_m = \frac{1}{2}\omega^m \int_0^1 \frac{\bar{E}_{m-1}(h-t)}{(m-1)!} f^{(m)}(x + \omega t)\, dt.$$

This summation formula is easy to establish by repeated integration by parts of the above integral. It is remarked in [**9**, p. 26] that this formula was known to Euler, for polynomial $f$ and without the remainder term. Also note that Lemma 1 turns into Taylor's formula with Lagrange's remainder term if we replace $h$ by $h/\omega$ and let $\omega$ approach zero.

To derive a convenient version of Lemma 1 for the applications we have in mind, we set $\omega = 1$ and impose further restrictions on $f$.

LEMMA 2. *Let $f$ be a function with $m$ continuous derivatives, defined on $t \geqslant x$. Suppose that $f^{(k)}(t) \to 0$ as $t \to \infty$ for all $k = 0, 1, \ldots, m$. Then for $0 \leqslant h \leqslant 1$*

$$\sum_{v=0}^{\infty} (-1)^v f(x + h + v) = \sum_{k=0}^{m-1} \frac{E_k(h)}{2k!} f^{(k)}(x) + R_m,$$

*where*

$$R_m = \frac{1}{2} \int_0^{\infty} \frac{\bar{E}_{m-1}(h-t)}{(m-1)!} f^{(m)}(x+t)\, dt.$$

**3. The Remainder for Gregory's Series.** The Euler numbers $E_n$ may also be defined by the generating function

$$\frac{2}{e^t + e^{-t}} = \sum_{n=0}^{\infty} E_n \frac{t^n}{n!}. \tag{3.1}$$

Comparing (3.1) with (2.1), we see that

$$E_n = 2^n E_n\left(\frac{1}{2}\right). \tag{3.2}$$

The phenomenon mentioned in the introduction is entirely explained by the next proposition—if we set $n = 500{,}000$. It is also clear that we will get similar patterns for $n = 10^m/2$ with any positive integer $m$.

PROPOSITION 1. *For positive integers n and M we have*

$$4\sum_{k=n}^{\infty} \frac{(-1)^k}{2k+1} = (-1)^n \sum_{k=0}^{M} \frac{2E_{2k}}{(2n)^{2k+1}} + R_1(M), \tag{3.3}$$

*where*

$$|R_1(M)| \leqslant \frac{2|E_{2M}|}{(2n)^{2M+1}}.$$

*Proof.* Apply Lemma 2 with $f(x) = 1/x$; then set $x = n$ and $h = 1/2$. We get

$$\sum_{v=0}^{\infty} \frac{(-1)^v}{n+v+1/2} = \sum_{k=0}^{m-1} \frac{E_k(1/2)}{2k!} \frac{(-1)^k k!}{n^{k+1}} + R_m, \tag{3.4}$$

with

$$R_m = \frac{1}{2}\int_0^{\infty} \frac{\bar{E}_{m-1}(h-t)}{(m-1)!} \frac{(-1)^m m!}{(x+t)^{m+1}} dt.$$

We multiply both sides of (3.4) by $2(-1)^n$. Then the left-hand side is seen to be identical with the left-hand side of (3.3). After replacing $m$ by $2M+1$ and taking into account (3.2) and the fact that odd-index Euler numbers vanish, we see that the first terms on the right-hand sides of (3.3) and (3.4) agree. To estimate the error term, we use the following inequality,

$$|E_{2M}(x)| \leqslant 2^{-2M}|E_{2M}| \quad \text{for } 0 \leqslant x \leqslant 1$$

(see, e.g., [**1**, p. 805]). Carrying out the integration now leads to the error estimate given in Proposition 1. □

**4. An Analogue For log 2.** Lemma 2 can also be used to derive a result similar to Proposition 1, concerning truncations of the series

$$\log 2 = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k}. \tag{4.1}$$

In this case the tangent numbers $T_n$ will play the role of the $E_n$ in Proposition 1. It

follows from the identity

$$\tan z = \frac{1}{i}\left(\frac{2e^{2iz}}{e^{2iz}+1} - 1\right)$$

together with (1.2) and (2.1) that

$$T_n = (-1)^n 2^n E_n(1) \tag{4.2}$$

as in [9, p. 28]. The $T_n$ can be computed using the recurrence relation $T_0 = 1$ and

$$\sum_{k=0}^{n} \binom{n}{k} 2^k T_{n-k} + T_n = 0 \quad \text{for } n \geqslant 1.$$

Other properties can be found, e.g., in [8] or [9, Ch. 2]:

PROPOSITION 2. *For positive integers n and M we have*

$$\sum_{k=n+1}^{\infty} \frac{(-1)^{k+1}}{k} = (-1)^{n+1}\left\{\frac{1}{2n} + \sum_{k=1}^{M} \frac{T_{2k-1}}{(2n)^{2k}}\right\} + R_2(M), \tag{4.3}$$

*where*

$$|R_2(M)| \leqslant \frac{|E_{2M}|}{(2n)^{2M+1}}.$$

*Proof.* We proceed as in the proof of Proposition 1. Here we take $x = n$ and $h = 1$. Using (4.2) and the fact that $T_0 = 1$ and $T_{2k} = 0$ for $k \geqslant 1$, we get the summation on the right-hand side of (4.3). The remainder term is estimated as in the proof of Proposition 1. □

Using Proposition 2 with $n = 10^m/2$ one again gets many more correct digits of $\log 2$ than is suggested by the error term of the Taylor series.

**5. Generalizations.** Proposition 1 and 2 can be extended easily in two different directions.

i). The well-known infinite series (see, e.g., [1, p. 807])

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)^{2n+1}} = \frac{|E_{2n}|}{2^{2n+2}(2n)!}\pi^{2n+1} \qquad (n = 0, 1, \dots), \tag{5.1}$$

and

$$\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k^{2n}} = (1 - 2^{1-2n})\zeta(2n)$$

$$= (2^{2n-1} - 1)\frac{|B_{2n}|}{(2n)!}\pi^{2n} \qquad (n = 1, 2, \dots) \tag{5.2}$$

can be considered as extensions of Gregory's series and of (4.1). These series admit exact analogues to Propositions 1 and 2; one only has to replace $f(x) = 1/x$ by $f(x) = x^{-(2n+1)}$, respectively $x^{-(2n)}$, in the proofs.

We note that the Euler-MacLaurin summation formula leads to similar results for

$$\sum_{k=1}^{\infty} k^{-2n} = \frac{|B_{2n}|2^{2n-1}}{(2n)!}\pi^{2n}, \tag{5.3}$$

where multiples of the Bernoulli numbers $B_{2n}$ take the place of the $E_n$ and $T_n$ in Propositions 1 and 2.

ii). A generalization of the Euler-MacLaurin and Boole summation formulas was derived by Berndt [3]. This can be applied to character analogues of the series (5.1)–(5.3). The roles of the $E_n$ and $T_n$ in Proposition 1 and 2 are then played by generalized Bernoulli numbers or by related numbers.

**6. Additional Comments.** The phenomenon observed in the introduction results from taking $N$ to be a power of ten; taking $N = 2 \cdot 10^m$ also leads to "clean" expressions. References [**1**], [**5**], [**6**], and [**9**] include the basic material on Bernoulli and Euler numbers, while [**8**] deals extensively with their calculation, and [**2**] describes an entertaining analogue of Pascal's triangle. Much on the calculation of pi and related matters may be found in [**4**]. Euler summation is treated in [**5**], [**6**], and [**9**], while Boole summation is treated in [**9**]. Related material on the computation and acceleration of alternating series is given in [**7**].

Added in Proof. A version of the phenomeon was observed by M. R. Powell and various explanations were offered (see *The Mathematical Gazette*, 66 (1982) 220–221, and 67(1983) 171–188).

REFERENCES

1. M. Abramowitz and I. Stegun, Handbook of Mathematical Functions, Dover, N.Y., 1964.
2. M. D. Atkinson, How to compute the series expansions of sec $x$ and tan $x$, *Amer. Math. Monthly*, 93 (1986) 387-388.
3. B. C. Berndt, Character analogues of the Poisson and Euler-MacLaurin summation formulas with applications, *J. Number Theory*, 7 (1975 413-445.
4. J. M. Borwein and P. B. Borwein, Pi and the AGM - A Study in Analytic Number Theory and Computational Complexity, Wiley, N.Y., 1987.
5. T. J. I'a Bromwich, An Introduction to the Theory of Infinite Series, 2nd ed., MacMillan, London, 1926.
6. R. L. Graham, D. E. Knuth, and O. Patashnik, Concrete Mathematics, Addison Wesley, Reading, Mass., 1989.
7. R. Johnsonbaugh, Summing an alternating series, this MONTHLY, 86 (1979) 637-648.
8. D. E. Knuth and T. J. Buckholtz, Computation of Tangent, Euler, and Bernoulli numbers, *Math. Comput.*, 21 (1967) 663-688.
9. N. Nörlund, *Vorlesungen über Differenzenrechnung*, Springer-Verlag, Berlin, 1924.
10. R. D. North, personal communications, 1988.
11. N. J. A. Sloane, A Handbook of Integer Sequences, Academic Press, New York, 1973.

# The William Lowell Putnam Mathematical Competition

LEONARD F. KLOSINSKI, *Santa Clara University*

G. L. ALEXANDERSON, *Santa Clara University*

LOREN C. LARSON, *St. Olaf College*

The following results of the forty-ninth William Lowell Putnam Mathematical Competition, held on December 3, 1988, have been determined in accordance with the governing regulations. This annual contest is supported by the William Lowell Putnam Prize Fund for the Promotion of Scholarship, left by Mrs. Putnam in memory of her husband, and is held under the auspices of the Mathematical Association of America.

The first prize, \$5,000, was awarded to the Department of Mathematics of Harvard University. The members of the winning team were: David J. Moews, Bjorn M. Poonen, and Constantin S. Teleman; each was awarded a prize of \$250.

The second prize, \$2,500, was awarded to the Department of Mathematics of Princeton University. The members of the winning team were: Daniel J. Bernstein, David J. Grabiner, and Matthew D. Mullin; each was awarded a prize of \$200.

The third prize, \$1,500, was awarded to the Department of Mathematics of Rice University. The members of the winning team were: Hubert L. Bray, Thomas M. Hyer, and John W. McIntosh; each was awarded a prize of \$150.

The fourth prize, \$1,000, was awarded to the Department of Mathematics of the University of Waterloo. The members of the winning team were: Frank M. D'Ippolito, Colin M. Springer, and Minh-Tue Vo; each was awarded a prize of \$100.

The fifth prize, \$500, was awarded to the Department of Mathematics of the California Institute of Technology. The members of the winning team were: William P. Cross, Robert G. Southworth, and Glenn P. Tesler; each was awarded a prize of \$50.

The five highest-ranking individual contestants, in alphabetical order, were David J. Grabiner, Princeton University; Jeremy A. Kahn, Harvard University; David J. Moews, Harvard University; Bjorn M. Poonen, Harvard University; and Ravi D. Vakil, University of Toronto. Each of these students was designated a Putnam Fellow by the Mathematical Association of America and awarded a prize of \$500 by the Putnam Prize Fund.

The next six highest-ranking individuals, in alphabetical order, were William P. Cross, California Institute of Technology; Serge Elnitsky, Carleton University; Karl M. Westerberg, Carnegie-Mellon University; Glen T. Whitney, Harvard University; Sihao Wu, Yale University; and Joshua R. Zucker, Stanford University. Each was awarded a prize of \$250.

The following teams, named in alphabetical order, received honorable mention: Brown University, with team members Peter H. Golde, Kevin S. McFarland, and David J. Morin; University of California, Berkeley, with team members I-Lin Kuo, Jordan Lampe, and David P. Moulton; Carnegie-Mellon University, with team members Petros I. Hadjicostas, Joseph G. Keane, and Karl M. Westerberg;

Stanford University, with team members John C. Loftin, John A. Overdeck, and Joshua R. Zucker; and Yale University, with team members Moses G. Klein, Robert S. Manning, and William N. Nelson.

Honorable mention was achieved by the following thirty-eight individuals named in alphabetical order: Thomas R. Amoth, Oregon State University; Tibor Beke, Armand Hammer United World College; Daniel J. Bernstein, Princeton University; David T. Blackston, Massachusetts Institute of Technology; Hubert L. Bray, Rice University; Jackson A. Bross, Massachusetts Institute of Technology; Timothy K. Callahan, University of Chicago; William Chen, Washington University, St. Louis; David Cook, Harvard University; Matthew M. Cook, University of Illinois, Urbana-Champaign; Peter H. Golde, Brown University; Thomas R. Hagedorn, Princeton University; Graydon H. Hazenberg, University of Waterloo; Thomas M. Hyer, Rice University; Joseph G. Keane, Carnegie-Mellon University; Moses G. Klein, Yale University; Matthew A. Klimesh, University of Michigan, Ann Arbor; Tal N. Kubo, Harvard University; John W. McIntosh, Rice University; Christopher J. Monsour, University of Maryland, College Park; David P. Moulton, University of California, Berkeley; Matthew D. Mullin, Princeton University; Du Nguyen, University of Ottawa; John A. Overdeck, Stanford University; David L. Petry, University of Oregon; Edward R. Ratner, California Institute of Technology; Raymond M. Sidney, Harvard University; Stephen A. Smith, University of Waterloo; Robert G. Southworth, California Institute of Technology; Colin M. Springer, University of Waterloo; Constantin S. Teleman, Harvard University; John S. Tillinghast, University of California, Davis; Minh-Tue Vo, University of Waterloo; Eric K. Wepsic, Harvard University; Christopher S. Wilson, Stanford University; David Bruce Wilson, Massachusetts Institute of Technology; John H. Woo, Harvard University; and Japheth L. M. Wood, Washington University, St. Louis.

The other individuals who achieved ranks among the top 104, in alphabetical order of their schools, were: Baylor University, Adrian Tanner; University of British Columbia, Wayne J. Broughton; Brown University, David J. Morin; California Institute of Technology, Ian Agol, Earl A. Hubbell, Russell A. Manning, Glenn P. Tesler; California Polytechnic State University, Daniel L. Krejsa; University of California, Berkeley, I-Lin Kuo, Jordan Lampe; University of California, Davis, Rudolf Von Bunau; University of Chicago, Andrew S. Yeh; Cornell University, Scott S. Benson; Emory University, Charles D. McDonell; Georgia Institute of Technology, Jeffrey W. Herrmann; Harvard University, Todd A. Brun, Duff G. Campbell, Leigh Chao, Roland B. Drier, Daniel D. Lee, David M. Maymudes, Michael D. Mitzenmacher, Daniel S. Sage, Michael E. Zieve; Hofstra University, Michael Cole; Iowa State University, Brad W. Michael; Knox College, Peter F. Schultz; Massachusetts Institute of Technology, Erik Ordentlich, Deniz Yuret; Michigan State University, Steven D. Fischer, Jacob R. Lorch; University of Michigan, Ann Arbor, Robert B. Doorenbos; University of Pennsylvania, Michael Albert; Princeton University, Emory F. Bunn, Timothy Y. Chow, David C. Fox, Rahul V. Pandharipande; Reed College, Nathaniel J. Thurston; University of Rochester, Daniel B. Finn; St. Olaf College, James S. Larson; Stanford University, Thomas H. Chung, John C. Loftin; Swarthmore College, Robert E. Marx; University of Texas, Austin, Douglas S. Hauge; University of Toronto, Edward J. Doolittle; Washington University, St. Louis, Peter S. Shawhan; University of Waterloo, Paulina Chin, Frank M. D'Ippolito, Michael Glaum, Simon H. Lee,

David N. C. Tse; University of Wisconsin, Madison, David G. Radcliffe; Yale University, Robert S. Manning, William M. Nelson; and Yeshiva University, Philip T. Reiss.

There were 2096 individual contestants from 360 colleges and universities in Canada and the United States in the competition of December 3, 1988. Teams were entered by 257 institutions.

The Questions Committee for the forty-ninth competition consisted of Abraham P. Hillman (Chair), Paul R. Halmos, and Gerald A. Heuer; they composed the problems listed below and were most prominent among those suggesting solutions.

## PROBLEMS

*Problem A-1*

Let $R$ be the region consisting of the points $(x, y)$ of the cartesian plane satisfying both $|x| - |y| \leqslant 1$ and $|y| \leqslant 1$. Sketch the region $R$ and find its area.

*Problem A-2*

A not uncommon calculus mistake is to believe that the product rule for derivatives says that $(fg)' = f'g'$. If $f(x) = e^{x^2}$, determine, with proof, whether there exists an open interval $(a, b)$ and a nonzero function $g$ defined on $(a, b)$ such that this wrong product rule is true for $x$ in $(a, b)$.

*Problem A-3*

Determine, with proof, the set of real numbers $x$ for which

$$\sum_{n=1}^{\infty}\left(\frac{1}{n}\csc\frac{1}{n} - 1\right)^{x}$$

converges.

*Problem A-4*

(a) If every point of the plane is painted one of three colors, do there necessarily exist two points of the same color exactly one inch apart?

(b) What if "three" is replaced by "nine"?

Justify your answers.

*Problem A-5*

Prove that there exists a *unique* function $f$ from the set $\mathbf{R}^+$ of positive real numbers to $\mathbf{R}^+$ such that

$$f(f(x)) = 6x - f(x) \quad \text{and} \quad f(x) > 0 \quad \text{for all } x > 0.$$

*Problem A-6*

If a linear transformation $\mathbf{A}$ on an $n$-dimensional vector space has $n + 1$ eigenvectors such that any $n$ of them are linearly independent, does it follow that $\mathbf{A}$ is a scalar multiple of the identity? Prove your answer.

*Problem B-1*

A *composite* (positive integer) is a product $ab$ with $a$ and $b$ not necessarily distinct integers in $\{2, 3, 4, \ldots\}$. Show that every composite is expressible as $xy + xz + yz + 1$, with $x$, $y$, and $z$ positive integers.

*Problem B-2*

Prove or disprove: If $x$ and $y$ are real numbers with $y \geqslant 0$ and $y(y+1) \leqslant (x+1)^2$, then $y(y-1) \leqslant x^2$.

*Problem B-3*

For every $n$ in the set $\mathbf{Z}^+ = \{1, 2, \ldots\}$ of positive integers, let $r_n$ be the minimum value of $|c - d\sqrt{3}|$ for all nonnegative integers $c$ and $d$ with $c + d = n$. Find, with proof, the smallest positive real number $g$ with $r_n \leqslant g$ for all $n$ in $\mathbf{Z}^+$.

*Problem B-4*

Prove that if $\sum_{n=1}^{\infty} a_n$ is a convergent series of positive real numbers, then so is $\sum_{n=1}^{\infty} (a_n)^{n/(n+1)}$.

*Problem B-5*

For positive integers $n$, let $\mathbf{M}_n$ be the $2n+1$ by $2n+1$ skew-symmetric matrix for which each entry in the first $n$ subdiagonals below the main diagonal is 1 and each of the remaining entries below the main diagonal is $-1$. Find, with proof, the rank of $\mathbf{M}_n$. (According to one definition the rank of a matrix is the largest $k$ such that there is a $k \times k$ submatrix with nonzero determinant.)

One may note that

$$\mathrm{M}_1 = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{pmatrix} \text{ and } \mathrm{M}_2 = \begin{pmatrix} 0 & -1 & -1 & 1 & 1 \\ 1 & 0 & -1 & -1 & 1 \\ 1 & 1 & 0 & -1 & -1 \\ -1 & 1 & 1 & 0 & -1 \\ -1 & -1 & 1 & 1 & 0 \end{pmatrix}.$$

*Problem B-6*

Prove that there exist an infinite number of ordered pairs $(a, b)$ of integers such that for every positive integer $t$ the number $at + b$ is a triangular number if and only if $t$ is a triangular number.

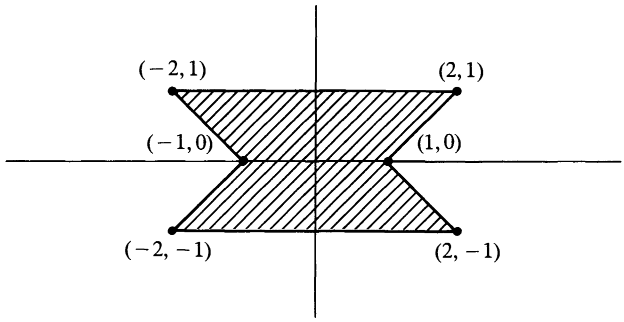
(The triangular numbers are the $t_n = n(n+1)/2$ with $n$ in $\{0, 1, 2, \ldots\}$.)

## SOLUTIONS

In the 12-tuples $(n_{10}, n_9, \ldots, n_0, n_{-1})$ following each problem number below, $n_i$ for $10 \geqslant i \geqslant 0$ is the number of students among the top 208 contestants achieving $i$ points for the problem and $n_{-1}$ is the number of those not submitting solutions.

*A-1.* $(202, 0, 0, 0, 0, 0, 0, 0, 4, 2, 0, 0)$

The part of $R$ in the first quadrant is bounded by $x = 0$, $y = 0$, $x - y = 1$, and $y = 1$. This part is a trapezoid with vertices $(0,0)$, $(1,0)$, $(2,1)$, and $(0,1)$ and area $3/2$. Since $(\pm x, \pm y)$ is in $R$ when $(x, y)$ is in $R$, the parts of $R$ in the other quadrants are obtained using symmetry about both axes, and consequently, the area of $R$ is 6.




*A-2.* (174, 1, 7, 1, 0, 0, 0, 3, 3, 0, 15, 4)

The function defined by $g(x) = e^x\sqrt{2x-1}$ has the property desired for $1/2 < a < x < b$ and $g(x) = e^x\sqrt{1-2x}$ has the property for $a < x < b < 1/2$.

To derive that result, consider the equation $(fg)' = f'g'$ and rewrite it in the successive forms

$$f'(x)g(x) + f(x)g'(x) = f'(x)g'(x),$$
$$\frac{g'(x)}{g(x)} = \frac{-f'(x)/f(x)}{1 - f'(x)/f(x)}.$$

If $f(x) = e^{x^2}$ then we have

$$\frac{g'(x)}{g(x)} = \frac{-2x}{1-2x}$$
$$\log|g(x)| = x + \tfrac{1}{2}\log|1-2x| + C,$$

where $C$ is an arbitrary constant. If $1/2 < a < x < b$, this has the form

$$g(x) = Ae^x\sqrt{2x-1},$$

where $A$ is an arbitrary positive real number. If $a < x < b < 1/2$, it has the form $g(x) = Ae^x\sqrt{1-2x}$.

*A-3.* (42, 17, 10, 0, 0, 0, 0, 6, 5, 3, 34, 91)

Let

$$a_n = \frac{1}{n}\csc\frac{1}{n} - 1.$$

Then

$$\begin{aligned} a_n &= \frac{1}{n\sin\frac{1}{n}} - 1 = \frac{1}{n\left(\frac{1}{n} - \frac{1}{6n^3} + \frac{1}{120n^5} - \cdots\right)} - 1 \\ &= \frac{1}{1 - \frac{1}{6n^2} + \frac{1}{120n^4} + \cdots} - 1 \\ &= 1 + \frac{1}{6n^2} + \frac{1}{n^2}g(n) - 1 = \frac{1}{n^2}\left(\frac{1}{6} + g(n)\right), \end{aligned}$$

where $g(n) \to 0$ as $n \to \infty$. Hence, there exist positive real numbers $c$, $d$, and $N$ such that

$$c\frac{1}{n^2} \leqslant a_n \leqslant d\frac{1}{n^2}, \quad \text{for } n > N.$$

Using the comparison and the $p$-test, one finds that $\Sigma a_n^x$ converges for $x > 1/2$ and diverges for $0 < x \leqslant 1/2$. But it is easy to see that the series also diverges for $x \leqslant 0$. Hence the answer is $\{x : x > 1/2\}$.

*A-4.* (45, 29, 2, 2, 1, 25, 11, 0, 3, 3, 25, 62)

(a) The answer is yes. For the proof let $A$ be an arbitrary point in the plane and let $ABC$ be an arbitrary equilateral triangle with side length 1 (where the units are inches, of course) that has $A$ as one of its vertices. If any two of $A$, $B$, and $C$ have the same color, the construction is finished. If not, let $A'$ be the point obtained by reflecting $A$ through the line $BC$. If $A'$ has the same color as either $B$ or $C$, the construction is finished. If not, then $A$ and $A'$ have the same color. Note that the distance between $A$ and $A'$ is $\sqrt{3}$, and that, in fact, any two points at distance $\sqrt{3}$ from one another can be obtained by making one of them a vertex of an equilateral triangle of side length 1 and then reflecting it through the side opposite it.

The result so obtained implies that, for any initial point $A$, either the reflected equilateral triangle argument finishes the desired construction for some $B$ and $C$, or else that every point at distance $\sqrt{3}$ from $A$ has the same color as $A$. The set of points such as $A'$, at distance $\sqrt{3}$ from $A$, is a circle of radius $\sqrt{3}$; any chord of length 1, of that circle, yields a pair of points of the same color exactly one inch apart.

(b) The answer is no. For the proof, pave the plane with squares whose common side length is chosen so that the diagonals are nearly 1 but not equal to 1; the diagonal length 0.9 will do. If that length is used, then the side length of each square is $0.9/\sqrt{2}$, which is somewhat greater than 0.63. Color one square with color #1, color the eight squares adjacent to it with colors #2–#9, and then repeat, throughout the plane, the coloring scheme of the large square (consisting of nine small squares) so obtained. (For present purposes it doesn't matter what consistent convention is followed for the boundaries of the squares; one possibility is to let the bottom and left boundaries of each square have the same color as the interior.)

The result is a nine-coloring of the plane in which no two points of the same color are exactly one inch apart. Indeed, for any point at all, the points of the same color are either within 0.9 inches from it or else farther than $2 \times .63 = 1.26$ inches.

*A-5.* (32, 8, 4, 0, 0, 0, 0, 1, 7, 96, 9, 51)

For arbitrary $x > 0$, let $a_0, a_1, a_2, \ldots$ be defined by $a_0 = x$ and $a_{n+1} = f(a_n)$. Then $a_{n+2} + a_{n+1} - 6a_n = 0$ for $n = 0, 1, 2, \ldots$. The characteristic roots of this difference equation are $-3$ and 2. Hence $a_n = (-3)^n c + 2^n k$ for some constants $c$ and $k$. As $a_{n+1} = f(a_n) > 0$ for all $n$, we must have $c = 0$ and so $f(x) = 2x$. This unique $f$ satisfies the conditions since it gives $f(f(x)) = f(2x) = 4x = 6x - f(x)$ and $2x > 0$ for $x > 0$.

*A-6.* (59, 14, 10, 7, 0, 0, 0, 0, 5, 0, 21, 92)

Yes, $\mathbf{A}$ must be a scalar multiple of the identity. Suppose that $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n+1}$ are eigenvectors of $\mathbf{A}$ such that any $n$ of them are linearly independent, with corre-

sponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_{n+1}$. Let $B_i = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n+1}\} \setminus \{\mathbf{x}_i\}$. Then $B_i$ is a set of $n$ linearly independent vectors in an $n$-dimensional vector space, so $B_i$ is a basis. With respect to $B_i$, the transformation $\mathbf{A}$ is represented by a diagonal matrix, $\operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{i-1}, \lambda_{i+1}, \ldots, \lambda_n)$. Thus, trace($\mathbf{A}$) $= S - \lambda_i$, where $S$ is the sum of the eigenvalues, $S = \lambda_1 + \lambda_2 + \cdots + \lambda_{n+1}$. But the trace of a linear transformation is independent of the basis chosen. Thus,

$$S - \lambda_i = S - \lambda_j \quad \text{for all } i, j$$

$$\lambda_i = \lambda_j \quad \text{for all } i, j$$

Let $\lambda$ be this common value. Then with respect to any of the bases $B_i$, $\mathbf{A}$ corresponds to $\operatorname{diag}(\lambda, \lambda, \ldots, \lambda)$, which is $\lambda$ times the identity.

*B-1.* (176, 25, 0, 0, 0, 0, 0, 0, 1, 1, 1, 4)
Letting $z = 1$, we have

$$xy + xz + yz + 1 = xy + x + y + 1 = (x + 1)(y + 1)$$

and this gives us all the composite positive integers when $x$ and $y$ range over all the positive integers.

*B-2.* (88, 38, 30, 0, 0, 0, 0, 0, 26, 4, 18, 4)
The desired conclusion is true for $0 \leqslant y \leqslant 1$, so suppose $y > 1$. If $(x + 1)^2 \geqslant y(y + 1)$ then we may assume that $x > 0$, and therefore $x \geqslant \sqrt{y(y + 1)} - 1 \geqslant \sqrt{y(y - 1)}$, and the result follows. (The last inequality follows from the easily verified fact that for positive numbers $a$ and $b$, $\sqrt{ab} + 1 \leqslant \sqrt{(a + 1)(b + 1)}$.)

*B-3.* (20, 16, 17, 2, 0, 0, 0, 0, 52, 33, 27, 41)
Let $g = (1 + \sqrt{3})/2$. Since for each fixed value of $n$ the sequence $n, n - 1 - \sqrt{3}, n - 2 - 2\sqrt{3}, \ldots, -n\sqrt{3}$ is an arithmetic progression with $-2g$ as common difference, there is a unique term $x_n$ in it with $-g < x_n < g$. Clearly $r_n = |x_n|$. Let $\varepsilon > 0$. By the pigeonhole principle, there exist $a$ and $b$ with $a \neq b$ and $|x_a - x_b| < \varepsilon$. Let $t = |a - b|$. In the sequence $r_t, r_{2t}, r_{3t}, \ldots$ there is an $r_{kt}$ such that $g - \varepsilon < r_{kt} \leqslant g$. Hence $g$ is the desired least upper bound of the $r_n$.

*B-4.* (17, 1, 0, 0, 0, 0, 0, 0, 3, 2, 73, 112)
Let $S = \{n : a_n^{n/n+1} < 2a_n\}$. If $n \notin S$, $a_n^{n/n+1} \geqslant 2a_n$, or equivalently $1/2 \geqslant a_n^{1-(n/n+1)} = a_n^{1/n+1}$, which is the same as $1/2^n \geqslant a_n^{n/n+1}$. It follows that

$$\sum_{n=1}^{\infty} a_n^{n/n+1} \leqslant \sum_{n \in S} a_n^{n/n+1} + \sum_{n \notin S} 1/2^n < \infty.$$

*B-5.* (9, 0, 10, 1, 0, 0, 0, 0, 0, 1, 2, 45, 140)
Let $u$ be a primitive $k$th root of 1, where $k = 2n + 1$. For $1 \leqslant i \leqslant k$, let $\mathbf{L}_i$ denote the column vector $(1, u^{i-1}, u^{2(i-1)}, \ldots, u^{(k-1)(i-1)})$. The $k$ by $k$ matrix whose $i$th column is $\mathbf{L}_i$ is a Vandermonde matrix, so the $\mathbf{L}_i$ are linearly independent over the complex numbers. For $1 \leqslant i \leqslant k$ we have $\mathbf{M}_n\mathbf{L}_i = c_i\mathbf{L}_i$, where $c_i$ is the scalar dot product of $\mathbf{L}_i$ with the first row of $\mathbf{M}_n$. Note that $c_i = 0$, but for $2 \leqslant i \leqslant k$, $c_i = -\sum_{j=2}^{n+1} u^{(j-1)(i-1)} + \sum_{j=n+2}^{k} u^{(j-1)(i-1)} \neq 0$. It follows that the vectors $\{c_2\mathbf{L}_2, \ldots, c_k\mathbf{L}_k\}$, and hence the vectors $\{\mathbf{M}_n\mathbf{L}_2, \ldots, \mathbf{M}_n\mathbf{L}_k\}$, are linearly independent. But $\mathbf{M}_n\mathbf{L}_1 = c_1\mathbf{L}_1 = 0$, so $\mathbf{M}_n$ has rank $2n$.

*B-6.* (38, 9, 4, 4, 24, 8, 0, 5, 5, 1, 17, 93)

**Solution 1**. It is easy to see that $t_{3n+1} = 1 + 9t_n$ and that $t_{3n} \equiv t_{3n+2} \equiv 0 \pmod 3$. This implies that $(9, 1)$ is one of our ordered pairs. If the numbers $a_m$ and $b_m$ are defined by

$$f(x) = 9x + 1, \qquad f(9x + 1) = a_2 x + b_2, \dots,$$
$$f(a_m x + b_m) = a_{m+1}x + b_{m+1},$$

an easy induction on $m$ shows that $(a_m, b_m)$ has the desired properties for $m = 2, 3, \dots$ .

**Solution 2**. We show that the ordered pairs $(8t_r + 1, t_r)$ have the desired properties. Let $T = \{0, 1, 3, 6, \dots\}$ be the set of triangular numbers and $Q = \{1, 9, 25, 49, \dots\}$ be the set of squares of odd integers. The equality $(2n + 1)^2 = 8((n^2 + n)/2) + 1$ implies that

$$t \text{ is in } T \text{ if and only if } 8t + 1 \text{ is in } Q. \qquad (*)$$

Let $t_r = r(r + 1)/2$ be in $T$ and $q = 8t_r + 1$.

For the "if" part, let $t$ be in $T$. Since $Q$ is closed under multiplication and $8t + 1$ is in $Q$ by $(*)$, we see that

$$q(8t + 1) = 8qt + q = 8qt + 8t_r + 1 = 8(qt + t_r) + 1$$

is in $Q$ and hence $qt + t_r$ is in $T$ by $(*)$. This proves the "if" part.

For the "only if" part, let $t$ be an integer and $qt + t_r$ be in $T$. Then

$$8(qt + t_r) + 1 = 8[(8t_r + 1)t + t_r] + 1 = (8t_r + 1)(8t + 1)$$

is in $Q$. Since $8t + 1$ is an integer and is the quotient of squares in $Q$, it follows that $8t + 1$ itself is in $Q$. Then $(*)$ tells us that $t$ is in $T$. This completes the proof.

# Geometry of Continued Fractions

M. C. IRWIN

MICHAEL CHARLES IRWIN wrote his Ph.D. thesis (Cambridge University, U.K., 1962) on "Embeddings of Polyhedral Manifolds," and his early papers were concerned with piecewise linear topology, but from about 1970 onwards his research was mainly on dynamical systems. His most recent publications included work on invariant paths for Anosov diffeomorphisms. He was the author of the book *Smooth Dynamical Systems* published by Academic Press in 1980. He taught at the University of Liverpool, U.K., from 1961 until his death on March 11, 1988.

The representation of real numbers by continued fractions dates back to Bombelli in the 16th century. In the 17th century Huygens used them in constructing a model of the solar system; he had to approximate the ratios of periods of planets by ratios of numbers of teeth on corresponding gear wheels, keeping the latter within reasonable bounds. The nice thing about the continued fraction process is that, being completely intrinsic, it brings out very strongly the personality of each individual number $\alpha$. The digits in the decimal expansion of $\alpha$ are much less revealing since they relate to the arbitrary choice of 10 as basis. On the other hand, the great defect of continued fractions is that it is virtually impossible to use them for even the simplest algebraic computation involving two or more numbers.

There are several books devoted entirely to the subject of continued fractions (e.g., **[1]**, **[2]**, **[3]**, **[5]**), and many books on number theory give an elementary introduction to the subject. The proofs are not difficult, but they are usually algebraic, and I find that when I read them I have a tendency to lose sight of where they are leading. The following geometrical treatment of the easiest results may be of some help to readers who, like myself, need a picture of what is going on. Harold M. Stark has already given such a treatment in his book **[4]**, but I think that the version below produces the main results with greater economy.

The idea of continued fractions comes from the observation that (i) any number between $1/3$ and $1/2$ (say) can be written as $1/(2$ plus a remainder), and (ii) the remainder, being between $1/(n+1)$ and $1/n$ for some $n$, is susceptible to the same treatment as the original number. More formally, we can express any real number $\alpha$ as the sum of an integer $[\alpha]$, the *integral part* of $\alpha$, and a number $\{\alpha\}$ with $0 \le \{\alpha\} < 1$, the *fractional part* of $\alpha$. If we define inductively $\alpha_0 = \alpha$ and $\alpha_n = 1/\{\alpha_{n-1}\}$ for $n \ge 1$, we obtain a sequence of integers $a_n = [\alpha_n]$, with $a_n \ge 1$ for $n \ge 1$. Of course the process terminates if $\{\alpha_n\} = 0$, and, in this case, $\alpha$ is rational and has the value

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots + \cfrac{}{a_n}}}} = \frac{p_n}{q_n}, \tag{1}$$

say, where $p_n/q_n$ is in its lowest terms. If $\{\alpha_n\} \neq 0$, then the rational number $p_n/q_n$ of (1) is intuitively some sort of approximation to $\alpha$, since we get it by ignoring the remainder in $1/(a_n$ plus a remainder). We write

$$p_n/q_n = [a_0; a_1, a_2, \dots, a_n].$$

It is called the *nth-order convergent* or *approximant* of $\alpha$. If $\alpha$ is irrational, then we find that $p_n/q_n \to \alpha$ as $n \to \infty$, and we write

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}} = [a_0; a_1, a_2, a_3, \dots, a_n, \dots].$$

Example 1.

$$\frac{2}{3} = [0; 1, 2] = \cfrac{1}{1 + \cfrac{1}{2}},$$

and

$$\frac{2}{3} = [0; 1, 1, 1] = \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1}}},$$

the continued fraction process for 2/3 giving the former expression. In general, we have the trivial equality

$$[a_0; a_1, \dots, a_{n-1}, a_n] = [a_0; a_1, \dots, a_{n-1}, a_n - 1, 1].$$

Example 2. The golden ratio $(1 + \sqrt{5})/2$ has continued fraction expansion $[1; 1, 1, \dots]$, since both the ratio and the continued fraction are positive numbers satisfying the quadratic equation derived from $x = 1 + 1/x$.

We begin by examining in what sense the convergents of $\alpha$ approximate $\alpha$. We say that a fraction $p/q$ $(q > 0)$ is a *best approximation* to $\alpha$ if, for all fractions $p'/q'$ with $0 < q' \leq q$, $|q\alpha - p| < |q'\alpha - p'|$ unless $q = q'$, $p = p'$. At first sight, it would be more natural to use the inequality $|\alpha - p/q| < |\alpha - p'/q'|$, which, written in the form

$$\frac{1}{q}|q\alpha - p| < \frac{1}{q'}|q'\alpha - p'|,$$

is clearly weaker. This would allow more fractions to be called best approximations, but the relation with continued fractions is more complicated (see **[2]**). Convergents of a number $\alpha$ are characterized as best approximations to $\alpha$ in the following way:

THEOREM 1. (i) *If $\{\alpha\} < 1/2$, then the best approximations to $\alpha$ are precisely the convergents $p_n/q_n$ for $n \geq 0$.*

(ii) *If* $\{\alpha\} > 1/2$, *then the best approximations to* $\alpha$ *are precisely the convergents* $p_n/q_n$ *for* $n \geq 1$.

*Note* (i). If $\{\alpha\} > 1/2$ then trivially the integral best approximation to $\alpha$ is $[\alpha] + 1$. Since $1 < 1/\{\alpha\} = \alpha_1 < 2$, $a_1 = 1$, and we have two integral convergents of $\alpha$, $p_0/q_0 = a_0 = [\alpha]$ and $p_1/q_1 = a_0 + 1/1 = [\alpha] + 1$, of which only the latter is a best approximation.

(ii) If $\{\alpha\} = 1/2$, then $\alpha$ has no integral best approximation, since $[\alpha]$ and $[\alpha] + 1$ are equally good. Thus the first convergent $p_0/q_0 = [\alpha]$ is not strictly speaking a best approximation. Of course $p_1/q_1 = \alpha$ is a best approximation.

*Proof of Theorem* 1. The outline of the proof is as follows. We associate geometrically best approximations to $\alpha$ with points in the plane, which we call *best approximation points*. We describe a geometrical construction which, when applied repeatedly, picks out in succession all best approximation points. Finally we compute the coordinates of the best approximation points to show that the best approximations to $\alpha$ are the given convergents. We shall also find that several further properties of convergents of $\alpha$ emerge immediately from this proof (see Corollaries 3 to 5 below).

*Notations and definitions.* Following Stark, we use the same notation $P = (b, a)$ for points in the plane and for vectors. The *lattice* generated by two vectors $P$ and $Q$ is the set of all points $mP + nQ$ where $m$ and $n$ are integers. The (closed) *positive* $(P, Q)$*-cone with vertex* $U$ is the set of points $U + aP + bQ$ for all numbers $a \geq 0$ and $b \geq 0$. The *distance from a point* $A$ *to a line* $l$ *in the direction* $P$ (not parallel to $l$) is the length of the line segment joining $A$ to the unique point $A + aP$ on $l$. The ratio

$$(\text{distance from } A \text{ to } l)/(\text{distance from } A' \text{ to } l)$$

is independent of the direction $P$ in which it is measured, so we are free to order distances from points to $l$ using any direction that suits us. See FIGURE 1 for an illustration of the above ideas.

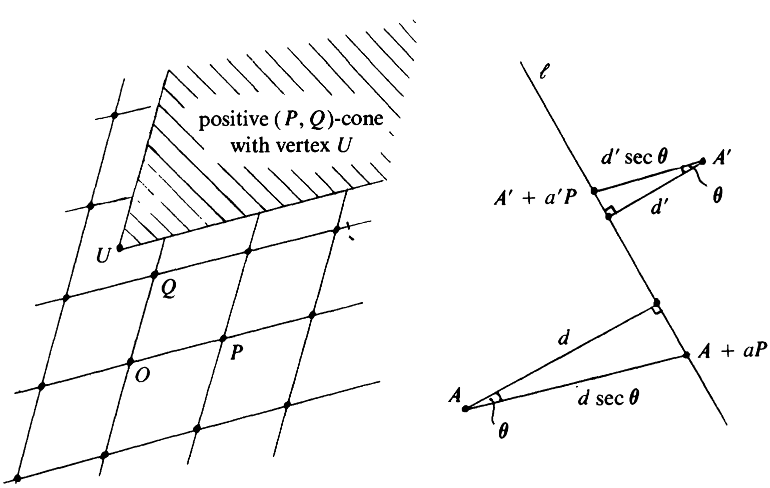



FIG. 1.

**Geometrical interpretation of best approximation.** Let $l$ be the line $y = \alpha x$. The geometrical interpretation of $p/q$ being a best approximation to $\alpha$ is that, amongst all points $(q', p')$ of the integer lattice $Z^2$ (the lattice generated by $(1,0)$ and $(0,1)$) with $x$-coordinate satisfying $0 < q' \leq q$, $(q, p)$ is uniquely the nearest to $l$. Here we use the fact that $|q'\alpha - p'|$ is the vertical distance from $(q', p')$ to $l$ (i.e., in the direction $(0,1)$). We call $(q, p)$ a *best approximation point* for $l$.

**The outpoint construction.** Consider the construction illustrated in FIG. 2: We are given a real number $\alpha$, and two integer lattice points $A$, $B$. We wish to construct the point $C$ in FIG. 2, which we will call the *outpoint* of $A$, $B$ and will denote by $C = C(A, B; \alpha)$.

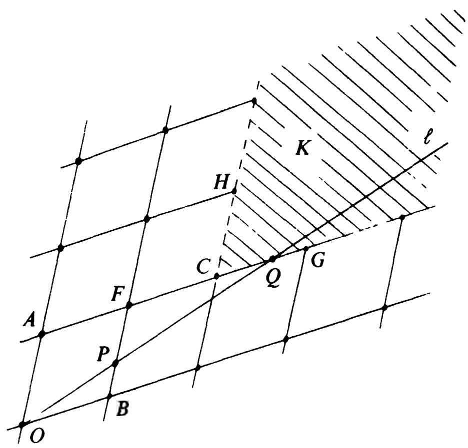



FIG. 2.

The outpoint construction is the following. Let $l$ be the line $y = \alpha\chi$. Suppose $l$ meets parallelogram $OBFA$ in the origin $O$, and in a point $P$ that is interior to the side $BF$. Let $l$ meet the extension of $AF$ in the point $Q$.

Then, $P = \theta A + B$ for some $\theta, 0 < \theta < 1$, and clearly $Q = A + (1/\theta)B$. Hence $Q$ lies on the lattice edge between $C = A + [1/\theta]B$ and $G = B + C$. This completes the construction of the outpoint $C = (A, B; \alpha)$. □

To what extent is $C$ a good lattice approximation to a point of $l$? Apart from 0 and $-C$, the reflection of $C$ in 0, $C$ is nearer to $l$ than is any lattice point outside (i) the positive $(A, B)$-cone $K$ with vertex at $C$ and (ii) the reflection of $K$ in 0. This is because the distance from any such lattice point to $l$ in the $B$-direction is at least a positive integral multiple of $|B|$, whereas the distance from $C$ to $l$ in this direction is $\varphi|B|$, and $\varphi < 1$. Moreover we can get away with a half-open cone for $K$, since any point on the $A$-edge of $K$ other than $C$ has distance greater than $|A|$ from $l$ in the $A$-direction. We also comment that, outside the two cones, the next nearest lattice points to $l$ after 0, $C$ and $-C$ are $B$ and $-B$, since these have distance $\theta|A|$ and $|B|$ in the $A$- and $B$-direction respectively.

**Repeating the outpoint construction.** We are going to iterate the outpoint construction, and in doing so, it is important to observe two facts. First, the lattice

generated by $A$ and $B$ is precisely the same as the lattice generated by $B$ and $C$ (because $C = A + [1/\theta]B$, $A = C - [1/\theta]B$ and $[1/\theta]$ is an integer). Second, at any vertex, the positive $(A, B)$-cone, open along the $A$-edge, contains the positive $(B, C)$-cone, open along the $B$-edge (because $C = A + [1/\theta]B$ and $[1/\theta] \geq 1$). Begin with a real number $\alpha$ and two integer lattice points $V_{-1} = (0,1)$ and $V_0 = (1, a_0)$, where $a_0 = [\alpha]$. Then for each $n = 1,2,3,\ldots$ let $V_n = C(V_{n-2}, V_{n-1}; \alpha)$. □

FIG. 3 illustrates the construction of $V_1, V_2, V_3$ when $\alpha = \sqrt{5} - 1 = [0;1,1,1,\ldots]$.

Trivially, if $\{\alpha\} < 1/2$, $p_0/q_0 = [\alpha]$ is a best approximation for $\alpha$. For $n \geq 1$, using the observations about lattices and positive cones, we can say that if $(q', p')$ is a point of $Z^2$ with $q' > 0$ and if $l$ is no further from $(q', p')$ than from $V_n$, then $(q', p')$ is in the positive $(V_{-1}, V_0)$-cone with vertex $V_n$, open along the $V_{-1}$-edge. Thus, by the geometrical interpretation, $p_n/q_n$ is a best approximation to $\alpha$. Also, by the next-nearest property of $B$ and $-B$, $l$ is closer to $V_{n-1}$ than to any point of $Z^2$ with $x$-coordinate between $q_{n-1}$ and $q_n$. Thus there is no best approximation $p/q$ with $q_{n-1} < q < q_n$. We deduce that there are no best approximations other than $p_n/q_n$, $n \geq 1$, and $p_0/q_0$ for $\{\alpha\} < 1/2$.

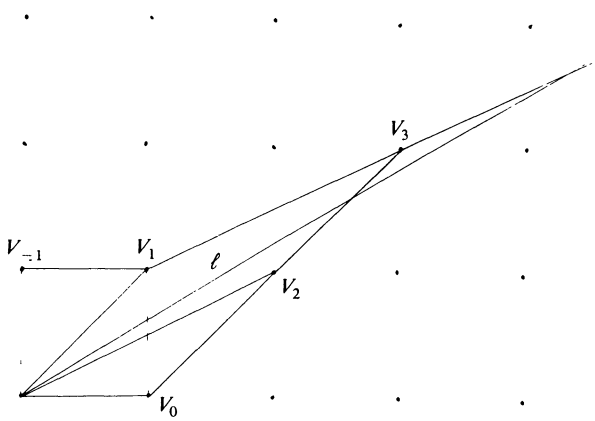



FIG. 3.

**Best approximations as convergents.** It remains to show that our $p_n/q_n$ satisfy (1). Let $l'$ be the line $y = (p_n/q_n)x$. First note the important fact that, for $0 \leq r \leq n$, the inductive sequence of constructions yields precisely the same points $V_r$ and integers $a_r$ for the line $l'$ as for the line $l$ (but, of course, unless $l = l'$ we get a new sequence of numbers $\theta_r'$ replacing $\theta_r$). This is because, for $0 \leq r < n$, $l'$ intersects the interior of the line segment joining $V_r$ to $V_r + V_{r-1}$. For certainly $l$ does so. Thus if $l'$ does not, then one of the lattice points $V_r$ or $V_r + V_{r-1}$ separates $l$ and $l'$ in the half-plane $x > 0$. Since either point has $x$-coordinate $\leq q_n$, it is nearer to $l$ than $V_n$ is, and this contradicts the fact that $p_n/q_n$ is a best approximation to $\alpha$.

Let $d_r'$ denote the vertical distance from $V_r$ to $l'$. Then $d_r'/d_{r-1}' = \theta_r'$ (This is because the distance ratio $CQ/OB$ equals $\varphi$ in the basic geometrical construction).

Hence,

$$d'_{r-2}/d'_{r-1} = 1/\theta'_{r-1} = a_r + \theta'_r = a_r + d'_r/d'_{r-1}$$

for $1 \le r \le n$, and so

$$d'_{r-1}/d'_{r-2} = 1/(a_r + d'_r/d'_{r-1}).$$

Now $d'_0/d'_{-1} = \{p_n/q_n\}$, and $d'_n = 0$. By substituting successively for the ratios $d'_1/d'_0, \dots, d'_{n-1}/d'_{n-2}$, we obtain $\{p_n/q_n\} = [0; a_1, \dots, a_n]$, and hence $p_n/q_n = [a_0; a_1, \dots, a_n]$. This completes the proof of Theorem 1.

We can easily deduce several basic properties of continued fractions from the above proof.

COROLLARY 2.

$$\begin{aligned} q_n &= a_n q_{n-1} + q_{n-2} \quad \textit{for } n \ge 1. \\ p_n &= a_n p_{n-1} + p_{n-2} \end{aligned}$$

*Proof.* Immediate from (2). In practice, one uses these relations to compute $q_n$ and $p_n$ inductively. The pair $q_n$, $p_n$ so obtained has H.C.F. 1, by the definition of best approximation. In the case of irrational $\alpha$, we deduce that the sequence $q_n$, $n \ge 0$, tends to $\infty$ at least as fast as the Fibonacci sequence $1, 1, 2, 3, 5, 8, \dots$ (since $q_n \ge q_{n-1} + q_{n-2}$). A cruder estimate is $q_n \ge 2^{n/2}$ for $n \ge 2$ (since $q_n \ge 0$, $q_n \ge q_{n-1}$ and thus $q_n \ge 2q_{n-2}$).

COROLLARY 3. *If $\alpha \ne p_n/q_n$, then $\alpha$ is between $p_{n-1}/q_{n-1}$ and $p_n/q_n$. The sequence $p_n/q_n$ increases for even n and decreases for odd n.*

*Proof.* Immediate from the geometrical construction.

COROLLARY 4. $p_{n-1}q_n - p_n q_{n-1} = (-1)^n$ *for* $n \ge 0$

*Proof.* The left-hand side is the oriented area of the parallelogram with vertices $0, V_n, V_n + V_{n-1}, V_{n-1}$. The result follows from the value $+1$ when $n = 0$, and the observation that, in the basic geometrical construction, the parallelograms $OBFA$ and $OCGB$ have the same base and height but opposite orientations.

Corollaries 2 and 3 give that, for $\alpha$ irrational,

$$|p_n/q_n - \alpha| < |p_n/q_n - p_{n-1}/q_{n-1}| = 1/q_n q_{n-1},$$

which shows that $p_n/q_n \to \alpha$ as $n \to \infty$, as we asserted in the introduction.

COROLLARY 5. $p_{n-2}q_n - p_n q_{n-2} = (-1)^{n-1}a_n$.

*Proof.* This follows from Corollary 4 with $n - 1$ replacing $n$, and the observation that, in the basic geometrical construction, the parallelograms $OBFA$ and $OCHA$ (where $H = A + C$) have the same base $OA$ and their heights are in the ration $OB/AC$, which is $1/[1/\theta]$.

Two numbers $\alpha$ and $\beta$ are said to be (*rationally*) *equivalent* if, for some integers $a, b, c, d$ with $ad - bc = \pm 1$,

$$\frac{a\alpha + b}{c\alpha + d} = \beta.$$

Geometrically, the second condition can be interpreted as a linear map $f$ with

matrix

$$\begin{pmatrix} d & c \\ b & a \end{pmatrix}$$

taking $(1, \alpha)$ to a multiple of $(1, \beta)$, or, equivalently, taking the line $y = \alpha x$ onto the line $y = \beta x$. The determinant condition $ad - bc = \pm 1$ is precisely the condition for $f$ to take the integer lattice $Z^2$ onto itself. It is also the condition for $f$ to induce a homeomorphism of the two-dimensional torus (which is obtained from the plane by identifying all points that differ by vectors in $Z^2$), so the idea is important, for example, when studying dynamical systems on the torus. Any rational number $\alpha = p/q$ (in its lowest terms) is equivalent to 0, since $mp + nq = 1$ for some integers $m$ and $n$, and thus

$$\begin{pmatrix} p & -q \\ n & m \end{pmatrix}\begin{pmatrix} q \\ p \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Hence any two rational numbers are equivalent. We finish this article with a well-known characterization of equivalence for irrational numbers. We shall say that the two continued fraction expansions $[a_0, a_1, a_2, \dots]$ and $[b_0; b_1, b_2, \dots]$ are *eventually identical* if for some $M \geq 0$ and $N \geq 0$, $a_{M+i} = b_{N+i}$ for all $i \geq 1$.

THEOREM 6. *Two irrational numbers are equivalent if and only if their continued fraction expansions are eventually identical.*

*Proof.* Let $\alpha$ and $\beta$ have continued fraction expansions as in the above definition of eventual identity. We prove that $\alpha$ is equivalent to $\gamma$, where $\gamma = [0; a_{M+1}, a_{M+2}, \dots]$. The similar result for $\beta$ then gives $\alpha$ equivalent to $\beta$. Let $l$ be the line $y = \alpha x$ and let $V_n$ be the associated best approximation points, as constructed in Theorem 1. Let $f$ be the linear map that takes $V_{M-1}$ and $V_M$ to $(0,1)$ and $(1,0)$ respectively. By Corollary 4, $f$ has determinant $\pm 1$. Let $l' = f(l)$ be the line $y = \gamma x$ and let $W_i = f(V_{M+i})$ for $i \geq -1$. Since $f$ is linear, it takes parallelograms to parallelograms, interior intersections to interior intersections, and linear sums $A + \theta B$ to $f(A) + \theta f(B)$. Thus it preserves the basic geometrical construction of Theorem 1. Hence it is clear that $W_i$, $i = 1, 2, \dots$ are the best approximation points for $l'$, and hence that $\gamma$ has continued fraction expansion $[0; a_{M+1}, a_{M+2}, \dots]$ as required.

Conversely, suppose that $\alpha = [a_0; a_1, a_2, \dots]$ is equivalent to $\beta = [b_0; b_1, b_2, \dots]$. By the preceding paragraph, we can assume that $a_0 = 0$. Let $f$ be a linear map taking $Z^2$ onto itself and the line $l$, $y = \alpha x$, onto the line $m$, $y = \beta x$. Again let $V_n$ be the best approximation points for $l$, and, this time, let $W_n = f(V_n)$. We can assume that the line $W_{-1}W_0$ intersects $m$ in some point $R$ lying in the half-plane $x > 0$ (otherwise, consider $-f$). By our comment of the preceding paragraph, $W_n$ is the sequence of points obtained by applying the inductive geometrical construction to the line $m$, beginning with the pair $W_{-1}$ and $W_0$, and $a_n$ is the corresponding sequence of integers obtained. Now let $d_n$ denote the distance from $W_n$ to $m$ in the direction of the line $W_{-1}W_0$. As in Theorem 1,

$$d_{n-2} = d_n + a_n d_{n-1},$$

whence $d_{n-1} \geq d_n$ and so $d_{n-2} \geq 2d_n$. Thus $d_n \to 0$ as $n \to \infty$. Now the line through $W_n$ parallel to $W_{-1}W_0$ intersects $m$ on the opposite side of $R$ to 0, so such intersections have a positive minimum distance from the half-plane $x \leq 0$. Thus, for

large enough $n$, $W_n$ has a positive $x$-coordinate. This implies that, for large enough $n$, the positive $(W_{n-2}, W_{n-1})$-cone at $W_n$ lies to the right of $W_n$. Hence, by the lattice approximation property of the basic geometrical construction, we recognise such $W_n$ as successive best approximation points of $m$. We deduce that the corresponding $a_n$ are successive terms in the continued fraction expansion of $\beta$, and this completes the proof of the theorem.

REFERENCES

1. W. B. Jones and W. J. Thron, Continued Fractions, Encyclopaedia of Mathematics and its Applications 11, Addison-Wesley, 1980.
2. A. Y. Khinchin, Continued Fractions, University of Chicago Press, 1964.
3. O. Perron, Die Lehre von den Kettenbrücke, Band I, Teubner, Stuttgart, 1954.
4. H. M. Stark, An Introduction to Number Theory, Markham, Chicago, 1970.
5. H. S. Wall, Analytic Theory of Continued Fractions, Van Nostrand, New York, 1948.

# The Editor's Corner: The White Screen Problem

HERBERT S. WILF

I bought my first home computer about nine years ago. I remember opening the carton, following the setup instructions minutely, and finally getting the thing working on my dining room table. Then I had to explain to myself why I had just spent all of that money, and the way to do that was to get it to do Something Interesting.

So I wrote a little Basic* program that would display a random walk on the screen. Precisely, it was the following. First, a pixel in the middle of the screen was lit up. Then one of the four directions N,E,S,W was selected uniformly at random, using the random number generator, and the walk proceeded one step in the chosen direction. That new pixel was lit up on the screen, and the process was repeated from the new point, etc.

The beauty of that was that it kept the computer working for a very long time and made a rather pretty, or anyway interesting, pattern on the screen, for the price of just a few program instructions. What sort of pattern emerged?

For a while, near the start of such a program, the walk is almost always quickly visiting pixels that it hasn't visited before, so one sees an irregular pattern that grows in the center of the screen. After a while, though, the walk will more often visit pixels that have previously been visited. Since they have already been lit up, and once lit up they are never turned off, the viewer sees no change on the screen.

Hence there are periods of time when the screen seems frozen, and then suddenly the walk will visit some new pixel in another corner of the pattern, and more of them will be lit up.

After quite a long while, when the screen is perhaps 95% illuminated, the growth process will have slowed down tremendously, and the viewer can safely go read *War and Peace* without missing any action. After a minor eternity, every cell will have been visited, the screen will be white, and the game will be over. Any mathematician who watches this will want to know how long, on the average, it will take before, for the first time, all pixels will have been visited.

To translate the question into more precise mathematical language, we consider a grid of $MN$ lattice points

$$G = \{(i,j) | 0 \leqslant i \leqslant M-1; 0 \leqslant j \leqslant N-1\}$$

and we regard them as being the vertices of a graph. The edges of the graph join $(i,j)$ to $(i,j+1), (i,j-1), (i-1,j), (i+1,j)$, in which the first coordinate is interpreted modulo $M$ and the second coordinate modulo $N$, because the screen was a 'wraparound' type where, for example, a point in the top row, if it walked up one unit, would appear on the bottom row (this graph is a lattice on a torus).

On that graph we take the random walk described above. The average epoch at which, for the first time, all vertices of the graph will have been visited at least once, we will call *the white screen time* of the graph $G$, and will denote it by $\omega(G)$.

*I'm still a closet Basic user.

Having gone so far, of course, we can define the notion on any connected graph $G$. Start at some vertex $v$. In general, having arrived at a vertex $w$, if $w$ has exactly $d(w)$ neighbors, choose one of them, with a priori probability $1/d(w)$, and go there next. The white screen time of the graph $G$ in this more general situation may depend not only on $G$ but also on the starting vertex $v$, and so we will then call it $\omega(G; v)$.

There's no need to stop there. The idea is perfectly natural not only on a connected graph, but on any irreducible Markov chain (i.e., one where every state can be reached from every other state in finitely many steps, with probability $> 0$). Then we would define the white screen time of the chain to be the expectation of the first epoch at which all states have been visited at least once. It would depend on the chain and on the initial state, in general.

Up to now we have followed a well-known mathematical strategy: if you can't solve it, generalize it. To continue with that strategy, let's solve some easier special cases of the more general problem. Here are three examples of finding the white screen time, each of which is well-known classical folklore.

**Example 1.** Consider the graph $P_n$, the $n$-path, that consists of $n$ points arranged in a line, with the $i$th point connected to the $(i+1)$st, for each $i = 1, \dots, n-1$ (no wraparound). If we begin a walk at point $j$, then what is $\omega(P_n; j)$?

To solve this we first consider a slightly different question. For each $j = 1, \dots, n$, let $x_j$ be the average number of steps that a walk will take if it starts at vertex $j$ of $P_n$, until for the first time it visits vertex $n$ (without necessarily having visited all vertices). Then clearly

$$x_j = \frac{1}{2}(x_{j+1} + x_{j-1}) + 1 \qquad (2 \leqslant j \leqslant n-1),$$

and we have the boundary conditions $x_n = 0$, $x_1 = x_2 + 1$.

It is easy to check that

$$x_j = n^2 - 2n + 2j - j^2 \qquad (j = 1, 2, \dots, n)$$

satisfies these equations, and since the solution is unique, that must be it.

In particular $x_1 = (n-1)^2$. But that means that we now have a partial answer to our original question, because

$$\omega(P_n; 1) = x_1 = (n-1)^2. \tag{1}$$

What about $\omega(P_n; j)$ for other $j$? If $y_j$ denotes that quantity then

$$y_j = \frac{1}{2}(y_{j+1} + y_{j-1}) + 1 \qquad (j = 2, 3, \dots, n-1)$$

along with the end conditions $y_n = y_1 = (n-1)^2$ by (1). This system has the solution

$$y_j = \omega(P_n; j) = \frac{5}{4}(n-1)^2 - \left(j - \frac{n+1}{2}\right)^2 \qquad (j = 1, 2, \dots, n),$$

so we now know the answers in the case of a path. ■

**Example 2.** Connect the two endpoints of an $n$-path and you have the $n$-cycle $C_n$. Here $\omega(C_n)$ is independent of the starting point.

To find it, fix an integer $j, 1 \leqslant j \leqslant n-1$. Consider the instant when, for the first time, our walk on the $n$-cycle has visited exactly $j$ vertices. Then these $j$ vertices are contiguous, and the walk is at an endpoint of that contiguous block at that moment. How long will it take before it visits a new point?

Consider the collection of $j$ already-visited points plus one new vertex adjoined at each end, as a path of $j+2$ vertices, with the walk now sitting at vertex 2. We are asking how long it will take before the walk visits *either* vertex 1 or vertex $j+2$. More generally, if we start at vertex $r$ of a $(j+2)$-path, how long does it take, on average, to reach an endpoint for the first time? Call that $z_r$. Then, as before,

$$z_r = \frac{1}{2}(z_{r-1}+1) + \frac{1}{2}(z_{r+1}+1) \qquad (2 \leqslant r \leqslant j+1),$$

with end conditions $z_1 = z_{j+2} = 0$, and with solution

$$z_r = (j+3)r - r^2 - (j+2) \qquad (r = 1, 2, \dots, j+2).$$

In particular, $z_2 = j$, which is the one we were interested in.

Therefore, for each $j = 1, 2, \dots, n-1$, if we have, for the first time, visited $j$ distinct points on the $n$-cycle, it takes an average of $j$ steps before a new point is encountered. Hence the white screen time of the $n$-cycle is (independent of the starting point and equal to)

$$\omega(C_n) = 1 + 2 + \cdots + (n-1) = \binom{n}{2}. \blacksquare \tag{2}$$

**Example 3.** Consider $n$ points, *every pair* of which are connected, and in which also each point is connected to itself. This is $\tilde{K}_n$, the 'complete graph with loops.' At each step, with probability $1/n$ the walk stays put, and otherwise it goes to some other vertex, all with equal probabilities.

The determination of the average number of steps required to visit all vertices in this case is called the *coupon collector's problem*, and is well known to probabilists. It can be restated as follows. If we repeatedly and independently select random integers from the set $1, 2, \dots, n$, how long will it take, on average, before we have for the first time selected every one of those $n$ integers at least once?

The answer is that an average of

$$\omega(\tilde{K}_n) = n\left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n}\right) \tag{3}$$

steps are needed. Indeed if we have already selected $j$ distinct integers then the probability that our next choice is a new one is $(n-j)/n$, so the average number of trials until we choose a new one is $n/(n-j)$, and the answer is the sum of $n/(n-j)$ from $j = 0$ to $n-1$, as claimed. ■

It turns out that problems of this type have been studied, under the name of *cover time*, by a number of authors (see [2] for the state of the art as of 1988). A few of the more recent results are that we know the order of magnitude of the white screen time of the $n$-cube [5], and that every connected graph $G$ of $n$ vertices has $\omega(G) > Cn \log n$ [1]. Further, for every connected $G$ of $n$ vertices and $E$ edges we have [3]

$$\omega(G) \leqslant 2nE. \tag{4}$$

Hence in particular $\omega(G) = O(n^3)$ for all $G$. The barbell graph $B_n$, consisting of two copies of the complete graph $K_n$ that are joined by an $n$-path, has $\omega(B_n) \sim Cn^3$, so the order of magnitude is best possible.

Here are a few things that we (or at any rate, I) don't know about this problem.

(a) Among all connected graphs $G$ of $n$ vertices which one has the largest $\max_v \omega(G; v)$? The smallest $\min_v \omega(G; v)$? The former is conjecturally a barbell. The latter is not the complete graph $K_n$ since if we adjoin a 'tail' to $K_{n-1}$ we can do better.

(b) It isn't immediately obvious that, given $G$, we can compute $\omega(G)$ in a finite amount of time. But there is a finite algorithm, based on inclusion-exclusion, that does so in time $Cn^a 2^n$ for graphs of $n$ vertices. Is the problem that hard really, or is it just that the algorithm isn't the best one?

(c) Is there a fast algorithm for computing $\omega(T)$ for a given *tree* $T$? There may well be, for Dr. David Aldous informs me that there is one for the closely related question of finding the mean time to cover and return to the starting point.

(d) I nearly forgot. What *is* the white screen time of the $M \times N$ toroidal lattice $T_{M,N}$? It has recently been shown by Zuckerman **[6]** that $\omega(T_{N,N})$ lies between $K_1 N^2 \log^2 N$ and $K_2 N^2 \log^2 N$. No wonder I had time to catch up on Tolstoy.

My thanks to Drs. David Aldous and Peter Matthews for a number of helpful comments on this writeup.

REFERENCES

1. David J. Aldous, Lower bounds for covering time for reversible Markov chains and random walks on graphs, *J. Theoretical Probability*, 2 (1989) 91–100.
2. ______, Bibliography: Random Walks on Graphs, 1988.
3. R. Aleliunas et al, Random walks, universal traversal sequences, and the complexity of maze traversal, *Proc. 20th IEEE Symposium on the Foundations of Computer Science*, (1979) 218–233.
4. P. Gerl, Random walks on graphs, in Probability Measures on Groups, v (H. Heyer, ed.), Springer Lecture Notes in Mathematics 1210, 1986.
5 Peter Claver Matthews, Some sample path properties of a random walk on the cube, *J. Theoretical Probability*, 2 (1989) 129–146.
6 David Zuckerman, Covering the 2-torus, preprint, 1989.

# LETTERS TO THE EDITOR

To the Editor:

It appears that Oscar A. Campoli, the author of "A Principal Ideal Domain which is not a Euclidean Domain", *Amer. Math. Monthly* 95 (1988), 868–871, is unaware of the paper with the same title by Jack C. Wilson, which appeared in *Math. Mag.* 46 (1973), 34–38.

The techniques used seem to me to be similar, any differences perhaps being explained by the fact that Wilson uses the standard definition for a Euclidean domain (the definition used by Campoli would not permit one to use the degree of a polynomial as a Euclidean norm for a polynomial ring over a field).

Yours sincerely

SHEILA OATES-WILLIAMS
University of Queensland
Australia

To the Editor:

Readers of D. H. Lehmer's article, "A New Approach to Bernoulli Polynomials" (this MONTHLY, December 1988), may be interested to know that the functional equation

$$\frac{1}{m}\sum_{k=0}^{m-1} B_n\left(x+\frac{k}{m}\right) = m^{-n}B_n(mx) \tag{1}$$

which forms the basis for Lehmer's article enters significantly in the modern theory of cyclotomic fields, as treated in *Cyclotomic Fields* by Serge Lang (Graduate Texts in Mathematics, vol. 59, Springer-Verlag, 1978). Lang calls (1) the "distribution relation," because it allows the $B_n$ to be interpreted in a way analogous to the theory of distributions on $\mathbb{R}^n$.

The distribution relation is also directly related to another important role of the Bernoulli polynomials, namely in the Euler-Maclaurin Summation Formula, one version of which may be formulated as follows:

For any $k \geqslant 1$ there is a bounded, $(k-2)$ times continuously differentiable function $J_k$ on $\mathbb{R}$, with period 1, such that for any $C^\infty$ function $f$ of compact support on $\mathbb{R}$, one has

$$\sum_{n=-\infty}^{\infty} f(n) - \int_{-\infty}^{\infty} f(t)\,dt = \int_{-\infty}^{\infty} J_k(t) f^{(k)}(t)\,dt. \tag{2}$$

Moreover the restriction of $J_k$ to the interval $[0,1]$ may be taken to be a polynomial

of degree $k$, such that

$$\int_0^1 J_k(t)\,dt = 0.$$

Subject to these normalizations, $J_k$ is unique.

The polynomials $J_k$ are simply multiples of the $B_k$:

$$B_k = (-1)^{k-1} k! J_k.$$

This result is easily established by induction on $k$, using the Fundamental Theorem of Calculus for $k = 1$, and integration by parts for $k > 1$. The distribution relation follows by an obvious change of variables $t \to mt$ in (2). Thus equation (2) provides a seventh, or perhaps a $6\frac{1}{2}$th, approach to the $B_n$.

Sincerely,

ROGER E. HOWE
Yale University

Editor:

I am, by profession, a computer software engineer, and, alas, only a very amateur historian of mathematics. As you are probably aware California has a significant Hispanic population. In the course of tutoring some high school students I was asked if there had ever been a famous Hispanic mathematician. I had some vague recollection and managed to track down a book I had read some years previously. Peter Beckmann, now an emeritus professor of engineering at Colorado, had written a history of pi, and mentioned a fifteenth century Spanish mathematician named Valmes. It seems that Valmes was alleged to have been burnt at the stake by Tomas de Torquemada in 1486 because the mathematician had asserted that he had solved the general quartic equation.

Besides providing me the example of a Spanish mathematician, the story aroused my interest. If Valmes had actually preceded Lodovico Ferrari's published solution of the quartic equation by some fifty years, it seemed only fair that Valmes get some credit for a feat that cost him his life. Even if Valmes' solution had been deficient I felt that mathematics owed a debt. But I could find no mention of Valmes in any history of mathematics, or in any history of the Spanish Inquisition.

I wrote to Professor Beckmann. He was very gracious and promptly sent a reply that his source for the story about Valmes was a Russian textbook titled *Tales About Mathematics* (Rasskazy O Matematike) by one Ivan I. Depman. Alas, there were no bibliographical references or footnotes about Valmes. The book was published in Leningrad by Gosdetizdat in 1954. The Soviet consulate here in San Francisco was kind enough to provide Gosdetizdat's address and I sent off a letter that had been professionally translated into Russian. I was optimistic that there might be some indication as to whether Depman, his editor or a student might still be alive. Unfortunately, there has been no reply. I also wrote to Henry Kamen in care of his

publisher. I am told that he is the foremost contemporary historian of the Spanish Inquisition. So far there has been no reply there either.

Is there a Russian mathematical journal whose editor and readers might be able to shed some light on Ivan Depman and his sources? Similarly, is there a Spanish mathematical journal that I could contact?

My last hope is that you might be able to print some or all of my letter and that one of my fellow readers might be able to point out that my problem has a known solution.

Many thanks for your time. I am

Sincerely yours,

Peter F. Zoll
P.O. Box 66
Port Costa, CA 94569

# NOTES

EDITED BY DENNIS DETURCK, DAVID J. HALLENBECK, AND ANITA E. SOLOW

## Fifty Years of Putnam Trivia*

JOSEPH A. GALLIAN
*Department of Mathematics and Statistics, University of Minnesota, Duluth, MN 55812*

Many of us eagerly await the announcement of the annual Putnam competition every March. We pore over the data looking for names of people we know and names that we recognize. We look to see how many Harvard entrants placed in the top 10 this year and how the team competition turned out. We are curious to see how previous winners did and what score it took to be a winner.

In 1988 there were two perfect scores. Two perfect scores! That is unbelievable. How many perfect scores have there been in the history of the competition? In 1987, and 1988, Harvard placed three in the top five. Incredible! Has any school ever had more? In 1988 Bjorn Poonen was a winner for the fourth time. How many times has that been done? In 1980 and 1982 Irwin Jungreis finished among the top ten while in 1985 and 1986 his brother Douglas was a Putnam Fellow. What a pair! Has one family ever produced two Putnam Fellows?

If you have ever asked yourself such questions then for you I have drawn up the following list of Putnam trivia questions. Enjoy.

1. Name two Putnam Fellows (top five finishers) who have won Nobel Prizes.
2. Name three Putnam Fellows who have won Fields Medals.
3. Name two Nobel Prize winners who have received Honorable Mention in the Putnam competition.
4. Has there ever been a team that had both a Nobel Laureate and a Fields Medalist?
5. Name a three-time Putnam Fellow who became President of the American Mathematical Society.
6. Name the first recipient of the Putnam Scholarship at Harvard.
7. Name the person who was a three-time winner of the USA Mathematical Olympiad and a three-time Putnam Fellow.
8. Besides Bjorn Poonen, who are the only other four-time Putnam Fellows?
9. How many contestants have been Putnam Fellows three times or more?
10. Which of the following have been Putnam Fellows: George Mackey, Felix Browder, R. G. Swan, Robin Hartshorne, Barry Simon, Alfred Hales?
11. Name the only person ever to have been designated a Putnam Fellow at two schools.
12. Six schools have had Putnam Fellows ten or more times. Name them.
13. Which of the following schools have never had a Putnam Fellow: Armstrong State College, Carleton College, College of St. Thomas, Fort Hays State

*This article is based on the Putnam competitions from 1938 to 1988.

10. All of them.
11. Arthur Rubin
12. Harvard 59, MIT 28, Cal Tech 17, Toronto 17, Berkeley 14, Princeton 11.
13. Carleton, Stanford
14. 3 (Harvard 1949, Harvard 1986, Harvard 1987)
15. 0
16. Harvard 14, Cal Tech 9, Toronto 4.
17. Second by Oberlin in 1972.
18. Four by Harvard, 1985–1988
19. Toronto, Queen's, Waterloo
20. Fourth
21. Polytechnic Institute and Brooklyn College
22. California, Davis
23. Stanford: one fourth place finish.
24. Each has won three times.
25. Michigan State University
26. Michigan State University

## Fixed Points of the Twisted Cyclic Shift Operator

LARRY W. CUSICK AND PETER TANNENBAUM
*Department of Mathematics, California State University, Fresno, CA 93740*

The cyclic shift operator $S(a_1,\dots,a_n)=(a_2,\dots,a_n,a_1)$ appears in many different mathematical contexts. It is, perhaps, the most natural representation of the cyclic group of order $n$, $\mathbb{Z}_n$. While attempting to solve a topological problem (see application 1) one of us has found useful a slight generalization of this operator. We felt that this new operator was interesting enough in its own right and that others might profit from its exposure. This generalization has also appeared elsewhere (see, for example, [3] and [4]).

Henceforth, we will suppose $A$ is a set with an involution. (An involution on a set $A$ is a function $\tau\colon A\to A$ that satisfies $\tau^2(a)=a$. Complex conjugation is a simple example of an involution. From now on we shall use $\bar{a}$ in place of $\tau(a)$, even when the involution is not complex conjugation.)

DEFINITION. *The twisted cyclic shift operator is the operator $T\colon A^n\to A^n$ given by*

$$T(a_1,\dots,a_n)=(a_2,\dots,a_n,\bar{a}_1).$$

The *exponent* of the twisted cycle shift operator is $2n$, i.e., $T^{2n}=\textit{identity}$, so we will find it convenient to think of $T$ as generating a cyclic group of order $2n$. In studying the dynamics of any operator one is often led to the fixed point sets of its powers,

$$\mathrm{Fix}(T^k)=\{\nu\in A^n\,|\,T^k(\nu)=\nu\}.$$

Our main result is a complete determination of the fixed point sets of powers of the twisted cyclic shift operator. Before we state the result we would like to introduce a little notation. The fixed points of the involution on $A$ will be denoted by $R(A)=\{a\in A\,|\,\bar{a}=a\}$ and if $u=(a_1,\dots,a_n)$ then $\bar{u}=(\bar{a}_1,\dots,\bar{a}_n)$.

10. All of them.
11. Arthur Rubin
12. Harvard 59, MIT 28, Cal Tech 17, Toronto 17, Berkeley 14, Princeton 11.
13. Carleton, Stanford
14. 3 (Harvard 1949, Harvard 1986, Harvard 1987)
15. 0
16. Harvard 14, Cal Tech 9, Toronto 4.
17. Second by Oberlin in 1972.
18. Four by Harvard, 1985–1988
19. Toronto, Queen's, Waterloo
20. Fourth
21. Polytechnic Institute and Brooklyn College
22. California, Davis
23. Stanford: one fourth place finish.
24. Each has won three times.
25. Michigan State University
26. Michigan State University

## Fixed Points of the Twisted Cyclic Shift Operator

LARRY W. CUSICK AND PETER TANNENBAUM
*Department of Mathematics, California State University, Fresno, CA 93740*

The cyclic shift operator $S(a_1,\dots,a_n) = (a_2,\dots,a_n,a_1)$ appears in many different mathematical contexts. It is, perhaps, the most natural representation of the cyclic group of order $n$, $\mathbb{Z}_n$. While attempting to solve a topological problem (see application 1) one of us has found useful a slight generalization of this operator. We felt that this new operator was interesting enough in its own right and that others might profit from its exposure. This generalization has also appeared elsewhere (see, for example, [3] and [4]).

Henceforth, we will suppose $A$ is a set with an involution. (An involution on a set $A$ is a function $\tau\colon A \to A$ that satisfies $\tau^2(a) = a$. Complex conjugation is a simple example of an involution. From now on we shall use $\bar{a}$ in place of $\tau(a)$, even when the involution is not complex conjugation.)

DEFINITION. *The twisted cyclic shift operator is the operator $T\colon A^n \to A^n$ given by*

$$T(a_1,\dots,a_n) = (a_2,\dots,a_n,\bar{a}_1).$$

The *exponent* of the twisted cycle shift operator is $2n$, i.e., $T^{2n} = \textit{identity}$, so we will find it convenient to think of $T$ as generating a cyclic group of order $2n$. In studying the dynamics of any operator one is often led to the fixed point sets of its powers,

$$\operatorname{Fix}(T^k) = \{\nu \in A^n \mid T^k(\nu) = \nu\}.$$

Our main result is a complete determination of the fixed point sets of powers of the twisted cyclic shift operator. Before we state the result we would like to introduce a little notation. The fixed points of the involution on $A$ will be denoted by $R(A) = \{a \in A \mid \bar{a} = a\}$ and if $u = (a_1,\dots,a_n)$ then $\bar{u} = (\bar{a}_1,\dots,\bar{a}_n)$.

MAIN THEOREM. *Let $k$ be any positive integer and $d = \gcd(k, 2n)$. Then*

$$\mathrm{Fix}(T^k) = \begin{cases} \{(u,\dots,u) \in A^n | u \in R(A)^d\} & \text{if } 2n/d \text{ is even}, \\ \{(u,\bar{u},\dots,u,\bar{u},u) \in A^n | u \in A^{d/2}\} & \text{if } 2n/d \text{ is odd}. \end{cases}$$

Before we prove the theorem we would like to illustrate the result with an example. (It would have been more honest to begin the article with the example since that would have better reflected the chronology of events. Indeed, the statement of the main theorem was inspired by the output of a computer program that calculated fixed point sets.)

EXAMPLE. Let $A = \{0,1\}$ with the involution $\bar{0} = 1$ and $\bar{1} = 0$ and $n = 9$. $\mathrm{Fix}(T^k)$ depends only on $d = \gcd(k, 18)$. We will restrict $k$ to the values 0 through 17. There are six possible values for $d$: the six divisors of 18, which are 1, 2, 3, 6, 9, and 18. The result is summarized in the following table:

| $d$ | $k$ | $\mathrm{Fix}(T^k)$ |
|---|---|---|
| 1 | 1, 5, 7, 11, 13, 17 | $\varnothing$ |
| 2 | 2, 4, 8, 10, 14, 16 | {(010101010), (101010101)} |
| 3 | 3, 15 | $\varnothing$ |
| 6 | 6, 12 | {(000111000), (001110001), (010101010), (100011100), (111000111), (110001110), (101010101), (011100011)} |
| 9 | 9 | $\varnothing$ |
| 18 | 0 | $A^9$ |

*Proof of the Main Theorem.* The first step is to note that $\mathrm{Fix}(T^k) = \mathrm{Fix}(T^d)$. This follows since $T^k$ and $T^d$ generate the same subgroup of $\langle T \rangle \approx \mathbb{Z}_{2n}$. (See, for example, [2]).

*Case* 1. $2n/d$ is even, so $d$ divides $n$. Write the elements of $A^n$ as vectors $v = (u_1, \dots, u_{n/d})$ where each $u_i$ is a vector in $A^d$. Then $T^d(v) = (u_2, \dots, u_{n/d}, \bar{u}_1)$ and $T^d(v) = v$ implies $u_1 = u_2 = \cdots = u_{n/d} = \bar{u}_1$. On the other hand, it should be obvious that any vector of the form $v = (u, \dots, u)$ with each $u \in R(A)^d$ is left fixed by $T^d$.

*Case* 2. $2n/d$ is odd, so $d$ is even. let $2p + 1 = 2n/d$. Write the elements of $A^n$ as vectors $v = (u_1, \dots, u_{2p+1})$ where each $u_i \in A^{d/2}$. Then $T^d(v) = (u_3, \dots, u_{2p+1}, \bar{u}_1, \bar{u}_2)$ and $T^d(v) = v$ implies $u_1 = u_3 = \cdots = u_{2p+1} = \bar{u}_2$ and $u_2 = u_4 = \cdots = u_{2p} = \bar{u}_1$. And clearly any vector in the form $v = (u, \bar{u}, \dots, \bar{u}, u)$ where $u \in A^{d/2}$ is left fixed by $T^d$. $\square$

An action by a group $G$ on a set $X$ is called free if for every $g \in G \setminus \{1\}$, $g(x) \neq x$ for all $x \in X$. An involution may be thought of as an action by the group $\mathbb{Z}_2$. Assume that the involution is free. In this case $R(A) = \varnothing$ so that $\mathrm{Fix}(T^k) = \varnothing$ whenever $2n/d$ is even. Consequently, if $n$ is a power of 2 then $\mathrm{Fix}(T^k)$ is empty for any positive $k$ less than $n$. The converse is true as well, for suppose $n = 2^q(2m+1)$ with $m > 1$ and let $k = 2^{q+1}$, then $2n/d = 2m + 1$. This gives us the following:

COROLLARY. *If the involution on $A$ is free then the resulting action of $\langle T \rangle \approx \mathbb{Z}_{2n}$ on $A^n$ is a free action if, and only if, $n$ is a power of 2.*

*Application* 1. There is a classical result in the topology of group actions that says the only nontrivial finite group that can act freely on an even dimensional sphere is $\mathbb{Z}_2$. (By an even dimensional sphere we mean any topological space homeomorphic to $S^{2n} = \{v \in \mathbb{R}^{2n+1} \mid \|v\| = 1\}$.) The proof of this result is both simple and instructive. The Euler characteristic, $\chi(X)$, plays a part in topology similar to that of cardinality in set theory. This is particularly striking when discussing group actions. For example, if a finite group, $G$, acts freely on a finite set, $X$, then the cardinalities of $X$ and the orbit set, $X/G$, are related to the order of $G$ by the formula $|X/G| \cdot |G| = |X|$. Now suppose $X$ is a "reasonable" compact topological space with a free action by $G$. (By "reasonable" we might mean that $X$ is a manifold or a polyhedron, and we assume the action preserves whatever structure we impose on $X$; in particular, the action should be continuous.) The Euler characteristics of $X$ and the orbit space $X/G$ now take the place of the cardinalities in the above formula:

$$\chi(X/G) \cdot |G| = \chi(X).$$

It is well known that when $X$ is an even dimensional sphere its Euler characteristic is 2. Now, we plug this into the above Euler characteristic formula to obtain $\chi(S^{2n}/G) \cdot |G| = 2$, from which we may conclude that $|G|$ divides 2. All that is left to do is to construct an actual free action of $\mathbb{Z}_2$ on $S^{2n}$. In fact, any sphere admits a free involution via the antipodal map, $\bar{v} = -v$, thus proving the result. We would like now to suggest a generalization. Suppose that $X$ is homeomorphic to a cartesian product $(S^{2n})^k$ and that $G$ is a finite group acting freely on $X$. Using the formula $\chi(A \times B) = \chi(A) \cdot \chi(B)$ and the fact $\chi(S^{2n}) = 2$, we conclude $\chi(X) = 2^k$. Using the same reasoning as above we get that $|G|$ must divide $2^k$. Note that the above condition on $G$ is necessary, but as we shall momentarily see, it is not sufficient. If $G$ is also assumed to be abelian then a further restriction on $G$ can be found. Any finite abelian 2-group is isomorphic to a direct product

$$G = \prod_{j=1}^{m} \mathbb{Z}_{2^{b_j}}$$

for some sequence of natural numbers $b_1, \dots, b_m$. In **[1]** and **[4]** it was proven that the above $G$ acts freely on $(S^{2n})^k$ if, and only if,

$$k \geqslant \sum_{j=1}^{m} 2^{b_j - 1}.$$

The "only if" part of the proof involves some representation theory which we will not go into here. The "if" part, though, is given by the above corollary. We need to show that if

$$k = \sum_{j=1}^{m} 2^{b_j - 1},$$

then we can construct a free action of $G$ on the cartesian product of $k$ copies of the $2n$-sphere. The construction is done in two steps. First, the twisted cyclic shift operator provides us with a free action of $\mathbb{Z}_{2^{b_j}}$ on each

$$(S^{2n})^{2^{b_j - 1}}$$

(use the antipodal map on the spheres for the involution). We now build up the action inductively using the following observation. If the groups $G_1$ and $G_2$ act freely on spaces $A_1$ and $A_2$ respectively, then the action of $G_1 \times G_2$ on $A_1 \times A_2$ given by $(g_1, g_2) \cdot (a_1, a_2) = (g_1 a_1, g_2 a_2)$ is also free. In this way we obtain a free action of $\mathbb{Z}_{2^{b_1}} \times \cdots \times \mathbb{Z}_{2^{b_n}}$ on

$$(S^{2n})^{2^{b_1-1}} \times \cdots \times (S^{2n})^{2^{b_n-1}}.$$

*Application* 2. This application originates in the design of digital circuits, an area where many interesting mathematical questions are submerged under heavy layers of nonmathematical terminology.

Here $A = \mathbb{Z}_2 = \{0,1\}$ with the natural involution $\bar{0} = 1$ and $\bar{1} = 0$. It is customary to write the elements of $A^n$ as $n$-bit strings rather than the more conventional vector notation and for the purposes of this example we will find it convenient to adhere to this custom.

In order to store an $n$-bit string (usually representing a specific piece of information such as a number or a character code) a device called a *register* is used. Registers are schematically represented as a collection of $n$ boxes each corresponding to an individual binary switch. The contents of the box describe the state of the switch (0 = off, 1 = on).

$$\boxed{1}\ \boxed{0}\ \boxed{0}\ \boxed{1}\ \boxed{1}\ \boxed{0}\ \boxed{1}\ \boxed{1}$$

FIG. 1. An 8-bit register containing the string 10011011.

Since registers by themselves are static and would be of little use except as a permanent storage device, logic is added in the form of circuitry and a discrete clock. The combination of storage device and logic circuitry gives rise to a device often known as a *counter* and the contents of the counter at a particular clock time $i$ is called the state of the counter at time $i$. Mathematically, a counter is just an operator $C$: $\mathbb{Z}_2^n \to \mathbb{Z}_2^n$ and the state of the counter at time $i$ is $C^i(a)$ where $a$ is the state of the counter at time 0.

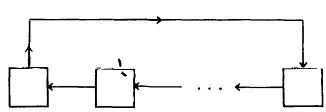


FIG. 2. The cyclic shift counter representing the cyclic shift operator $S$.

The sequence of binary strings generated by the consecutive states of a counter is called a *count sequence*. Count sequences are used to control sequences of operations and/or commands and in a rather simplistic sense a program is essentially a

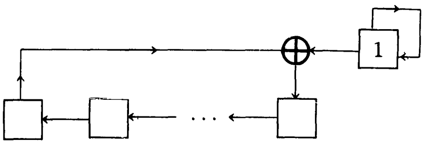



FIG. 3. The twisted cyclic shift counter or "Moebius counter" representing the operator $T$. The binary adder $\oplus$ performs the involution by adding the constant bit 1.

conglomeration of complex count sequences. The name of the game in logic design is to design a counter that generates a given count sequence in a way that minimizes the complexity of the circuitry.

One of the best known and widely used count sequences is a *Gray code sequence*. In a Gray code sequence the state of the counter at clock time $(i + 1)$ differs from the state of the counter at clock time $i$ in a single bit position. For obvious reasons, Gray code sequences are known as *minimal change sequences*. Although it is possible to generate a Gray code sequence of $2^n$ states using an $n$-bit counter the logic circuitry needed to implement it is quite complicated [**3**]. A much simpler design can be obtained by using an $(n + 1)$-bit twisted cyclic shift counter as illustrated in FIGURE 4.

| Clock Time | State |
|---|---|
| 0 | $0000 = 0$ |
| 1 | $0001 = T(0)$ |
| 2 | $0011 = T^2(0)$ |
| 3 | $0111 = T^3(0)$ |
| 4 | $1111 = T^4(0)$ |
| 5 | $1110 = T^5(0)$ |
| 6 | $1100 = T^6(0)$ |
| 7 | $1000 = T^7(0)$ |

FIG. 4. An eight state Gray code sequence generated by a 4-bit twisted cyclic shift counter.

One of the questions raised by this paper can be phrased as follows: given that the counter is in an initial state $a$, what is the longest possible Gray code sequence that it can generate? Using the terminology of our paper this is equivalent to asking what is the smallest power of the operator $T$ which fixes $a$? The answer to this question is given by the main theorem.

REFERENCES

1. L. W. Cusick, Finite abelian groups that can act freely on $(S^{2n})^k$, to appear in the *Illinois Journal of Mathematics*.
2. I. N. Herstein, Abstract Algebra, Macmillan, 1985.
3. T. R. Blakeslee, Digital Design with Standard MSI and LSI, Wiley, 1975.
4. M. Hoffman, Free actions of abelian groups on a cartesian power of an even sphere, to appear in the *Canadian Mathematical Bulletin*.

# Fubini's Theorem for Null Sets

ERIC K. VAN DOUWEN†
*Mathematics Department, North Texas State University, Denton, TX 76203*

For a subset $A$ of the plane $\mathbb{R}^2$ and for $x \in \mathbb{R}$ the set

$$A_x = \{ y \in \mathbb{R} : \langle x, y \rangle \in A \}$$

is called the **x-section** of $A$. We let $\lambda$ denote ($\lambda$inear) Lebesgue measure in $\mathbb{R}$ and let $\pi$ denote ($\pi$lanar) Lebesgue measure in $\mathbb{R}^2$.

Fubini's theorem for null sets is:

THEOREM A. *For a measurable subset $A$ of $\mathbb{R}^2$ the following are equivalent*:

(a) *$A$ is a null set of $\mathbb{R}^2$*;

(b) *$A_x$ is a null set for almost all $x \in \mathbb{R}$, i.e., the exceptional set*

$$E(A) = \{ x \in \mathbb{R} : A_x \textit{ is not a null set of } \mathbb{R} \}$$

*is a null set of $\mathbb{R}$.*

Oxtoby gives a simple proof of $(a) \Rightarrow (b)$, [**1**, p. 53]. Concerning the reverse implication he states that "[t]he usual proof depends on properties of the Lebesgue integral" [**1**, p. 54]. The purpose of this note is to give an unusual proof of the reverse implication that is quite in the spirit of [**1**].

Fubini's theorem for null sets does not give much information about the exceptional set $E(A)$ if $A$ is a measurable set that is not a null set. It leaves open the possibility that $E(A)$ has cardinality less than $\mathfrak{c}$ (this requires the negation of the Continuum Hypothesis, of course). Our method of proof allows us to prove "Fubini's Theorem for Sets of Positive Measure":

THEOREM B. *For a measurable subset $A$ of $\mathbb{R}^2$ the following are equivalent*:

(a) *$A$ is not a null set*;

(b) *there are $\gamma > 0$ and a measurable $G \subseteq \mathbb{R}$ of positive measure such that $\lambda(A_x) > \gamma$ for all $x \in G$.*

As an application of Theorem B we give a variation of an example of Sierpiński, [**2**], that "measurable" is essential in Theorems A and B: There is a subset $A$ of $\mathbb{R}^2$ that is not measurable, yet no vertical line in $\mathbb{R}^2$ contains more than one point of $A$. A similar example is also constructed by Oxtoby, [**1**, p. 55], who used $CH$ to keep things simple. Since we have Theorem B available, we can do a simple construction without using $CH$.

**1. Compact sets in products.** The key to our proof is the following simple result about compact sets in products:

*Fact.* Let $A \subseteq \mathbb{R}^2$ be compact, let $x \in \mathbb{R}$ and let $T$ be open in $\mathbb{R}$ with $A_x \subseteq T$. Then $x$ has a neighborhood $U$ such that

$$A \cap (U \times \mathbb{R}) \subseteq U \times T, \quad \text{i.e., such that } A_y \subseteq T \text{ for all } y \in U.$$

†Professor Eric K. van Douwen died in his sleep July 28, 1987. He was 41 years old.

To prove the fact note that

$$\mathscr{F} = \{A \cap ([a, b] \times (\mathbb{R} \setminus T)) : a < x < b\}$$

is a collection of closed subsets of the compact set $A$ with

$$\bigcap \mathscr{F} = A \cap (\{x\} \times (\mathbb{R} \setminus T)) = \{x\} \times (A_x \cap (\mathbb{R} \setminus T)) = \varnothing.$$

Since $\mathscr{F}$ is closed under finite intersection, it follows that $\varnothing \in \mathscr{F}$; hence there are $a, b \in \mathbb{R}$ with $a < x < b$ such that $A \cap ([a, b] \times \mathbb{R}) \subseteq [a, b] \times T$.

This proof works in any product $X \times Y$ with $X$ Hausdorff. Therefore, the two Fubini Theorems we prove also hold in a more general context. We leave it to the reader to elaborate.

**2. Proof of Fubini's Theorem for Null Sets.** Since $(a) \Rightarrow (b)$ has an easy proof, [**1**, p. 53], we only have to prove $(b) \Rightarrow (a)$. We first reduce the proof to the special case of compact sets with empty exceptional set: Let $A$ be a measurable subset of $\mathbb{R}^2$ whose exceptional set $E(A)$ is a null set. Since $E(A) \times \mathbb{R}$ is a null set, we prove $A$ is a null set if we prove $A \setminus (E(A) \times \mathbb{R})$ is a null set. To this end, it suffices to prove for each compact $K \subseteq A \setminus (E(A) \times \mathbb{R})$ that $K$ is a null set. As $E(S) \subseteq E(T)$ whenever $S \subseteq T \subseteq \mathbb{R}^2$, this completes the reduction. We may assume without loss of generality that $E(A) = \varnothing$ and that $A$ is compact.

Since $A$ is compact, so is its projection on the $x$-axis, i.e., so is

$$L = \{x \in \mathbb{R} : \langle x, y \rangle \in A \text{ for some } y \in \mathbb{R}\}.$$

Consider an arbitrary $\delta > 0$; we prove $\pi(A) < \delta$. As $L$ is compact, there is an open $S$ in $\mathbb{R}$ with $L \subseteq S$ such that $\lambda(S) < \infty$. Let $\varepsilon > 0$ be such that $\lambda(S) \cdot \varepsilon \leq \delta$. Call $U \subseteq \mathbb{R}$ $\varepsilon$-good if there is an open set $V$ in $\mathbb{R}$ with $\lambda(V) < \epsilon$ such that

$$A \cap (U \times \mathbb{R}) \subseteq U \times V,$$

and let $\mathscr{G}$ be the collection of $\varepsilon$-good open sets $G$ in $\mathbb{R}$ with $G \subseteq S$. Since $E(A) = \varnothing$, the Fact of Section 2 implies that $\mathscr{G}$ covers $L$. As $L$ is compact, it follows that $\mathscr{G}$ has a finite subcover, which we enumerate as $\langle G_k : k = 0, \dots, n \rangle$. From this finite open cover of $L$ by $\varepsilon$-good open sets we produce a pairwise disjoint cover $\langle S_k : k = 0, \dots, n \rangle$ of $L$ by $\varepsilon$-good measurable subsets of $\mathbb{R}$: Define

$$S_k = G_k \setminus \bigcup_{i<k} G_i, \quad \text{for } k = 0, \dots, n.$$

Each $S_k$ is $\varepsilon$-good since it is a subset of the $\varepsilon$-good set $G_k$. Also, for each $x \in L$, if $m$ is minimal with $x \in G_m$, then $x \in S_m$. Hence the $S_k$'s cover $L$.

For each $k = 0, \dots, n$ we can find an open $V_k$ in $\mathbb{R}$ with $\lambda(V_k) < \varepsilon$ such that

$$A \cap (S_k \times \mathbb{R}) \subseteq S_k \times V_k.$$

Since $\langle S_k : k = 0, \dots, n \rangle$ is pairwise disjoint and satisfies

$$L \subseteq \bigcup_{k=0}^{n} S_k \subseteq S,$$

since it has a countable base.) Hence there is a well-order $\prec$ on $\mathscr{P}$ such that every member of $\mathscr{P}$ has fewer than $\mathfrak{c}$ predecessors.

We construct $A$ by choosing $a_P = \langle x_P, y_P \rangle \in P$ for $P \in \mathscr{P}$ as follows: Let $Q \in \mathscr{P}$, and assume $a_P$ chosen for $P \prec Q$. Because of (b) we can choose $x_Q \in \mathbb{R} \setminus \{x_P : P \prec Q\}$ such that the vertical line $\{x_Q\} \times \mathbb{R}$ intersects $Q$, and next choose $y_Q \in \mathbb{R}$ such that $a_Q = \langle x_Q, y_Q \rangle \in Q$. Obviously no vertical line contains more than one point of the resulting set $A = \{a_P : P \in \mathscr{P}\}$.

Of course (b) also holds with "$|P_x| = \mathfrak{c}$" instead of "$P_x \neq \varnothing$." The reader may use this version of (b) to construct a subset $R^2$ which contains exactly two points of every line in $\mathbb{R}^2$, yet which is not a null set since, again, it meets every member of $\mathscr{P}$.

REFERENCES

1. J. C. Oxtoby, Measure and Category, Springer, Berlin, 1971.
2. W. Sierpiński, Sur une problème concernant les ensembles mesurables superficiellement, *Fund. Math.*, 1(1920) 112–115.

# An Optimization Problem

RICHARD BASSEIN
*Department of Mathematics and Computer Science, Mills College, Oakland, CA 94613*

Some years ago, I played with a computer game called "Lunar Lander," of which there are several forms in existence. The one I used simulated the following situation. You are piloting a lunar excursion module which has a given amount of fuel and is hurtling towards the surface of the moon. You must decide when and how hard to fire your retrorockets to slow your vehicle to a gentle landing. After much experimentation and consultation with other "simu-astronauts," I found that the best strategy was to wait as long as possible before firing the rockets, then blast away with the maximum force allowed. (That strategy is, in fact, an example of the so called "bang-bang" principle of optimal control; advanced treatments using functional analysis may be found in the references.) This article presents an elementary proof that such a strategy is indeed the best.

To make the central ideas most apparent, let's first consider a simple model in which the change in weight of the remaining fuel is negligible compared to the total weight of the vehicle. Let

$$\begin{aligned} x(t) &= \text{the height of the vehicle above the surface of the moon} \\ v(t) &= \text{the downward velocity of the vehicle; thus } v(t) = -x'(t) \\ f(t) &= \text{the cumulative amount of fuel consumed; thus } f(0) = 0 \\ g &= \text{the acceleration of gravity near the moon's surface} \\ T &= \text{the time at which the vehicle lands; thus } x(T) = 0. \end{aligned}$$

Our goal is to minimize $f(T)$ while making $v(T) = 0$, for a gentle landing.

Let us assume that the force applied by the retrorockets is proportional to the rate at which fuel is consumed. If $\underline{c}$ is the constant of proportionality and $\underline{m}$ is the

since it has a countable base.) Hence there is a well-order $\prec$ on $\mathscr{P}$ such that every member of $\mathscr{P}$ has fewer than $\mathfrak{c}$ predecessors.

We construct $A$ by choosing $a_P = \langle x_P, y_P \rangle \in P$ for $P \in \mathscr{P}$ as follows: Let $Q \in \mathscr{P}$, and assume $a_P$ chosen for $P \prec Q$. Because of (b) we can choose $x_Q \in \mathbb{R} \setminus \{x_P : P \prec Q\}$ such that the vertical line $\{x_Q\} \times \mathbb{R}$ intersects $Q$, and next choose $y_Q \in \mathbb{R}$ such that $a_Q = \langle x_Q, y_Q \rangle \in Q$. Obviously no vertical line contains more than one point of the resulting set $A = \{a_P : P \in \mathscr{P}\}$.

Of course (b) also holds with "$|P_x| = \mathfrak{c}$" instead of "$P_x \neq \varnothing$." The reader may use this version of (b) to construct a subset $R^2$ which contains exactly two points of every line in $\mathbb{R}^2$, yet which is not a null set since, again, it meets every member of $\mathscr{P}$.

REFERENCES

1. J. C. Oxtoby, Measure and Category, Springer, Berlin, 1971.
2. W. Sierpiński, Sur une problème concernant les ensembles mesurables superficiellement, *Fund. Math.*, 1(1920) 112–115.

# An Optimization Problem

RICHARD BASSEIN
*Department of Mathematics and Computer Science, Mills College, Oakland, CA 94613*

Some years ago, I played with a computer game called "Lunar Lander," of which there are several forms in existence. The one I used simulated the following situation. You are piloting a lunar excursion module which has a given amount of fuel and is hurtling towards the surface of the moon. You must decide when and how hard to fire your retrorockets to slow your vehicle to a gentle landing. After much experimentation and consultation with other "simu-astronauts," I found that the best strategy was to wait as long as possible before firing the rockets, then blast away with the maximum force allowed. (That strategy is, in fact, an example of the so called "bang-bang" principle of optimal control; advanced treatments using functional analysis may be found in the references.) This article presents an elementary proof that such a strategy is indeed the best.

To make the central ideas most apparent, let's first consider a simple model in which the change in weight of the remaining fuel is negligible compared to the total weight of the vehicle. Let

$$\begin{aligned} x(t) &= \text{the height of the vehicle above the surface of the moon} \\ v(t) &= \text{the downward velocity of the vehicle; thus } v(t) = -x'(t) \\ f(t) &= \text{the cumulative amount of fuel consumed; thus } f(0) = 0 \\ g &= \text{the acceleration of gravity near the moon's surface} \\ T &= \text{the time at which the vehicle lands; thus } x(T) = 0. \end{aligned}$$

Our goal is to minimize $f(T)$ while making $v(T) = 0$, for a gentle landing.

Let us assume that the force applied by the retrorockets is proportional to the rate at which fuel is consumed. If $c$ is the constant of proportionality and $m$ is the

mass of the vehicle (including fuel) then Newton's Law says the following about the downward velocity $v(t)$.

$$mv'(t) = mg - cf'(t) \tag{1}$$

By choosing the appropriate units for $f(t)$ we can make $c = m$ and this equation takes the following simpler form.

$$v'(t) = g - f'(t)$$

The use of $f'(t)$ when the force, and therefore the rate of fuel consumption, is changed deserves some discussion. While in reality such changes in $f'(t)$ may be rapid but continuous, it will be simpler, and not terribly inaccurate, to model them as discontinuous. The problem is that $f'(t)$ will be undefined at such moments. The easiest way to deal with this difficulty is to assume that $f(t)$ is continuous and piecewise differentiable and integrate the above equation.

Integrating $v'(t) + f'(t) = g$ gives $v(t) + f(t) = C + gt$, for some constant $C$. If the initial velocity of the vehicle is $v_0 = v(0)$, we get $C = v_0$ since $f(0) = 0$. Thus,

$$v(t) + f(t) = v_0 + gt \tag{2}$$

If $v(T) = 0$, then $f(T) = v_0 + gT$. Our first observation is that to minimize $f(T)$, we need to minimize $T$.

Fig. 1 illustrates equation 2 geometrically.

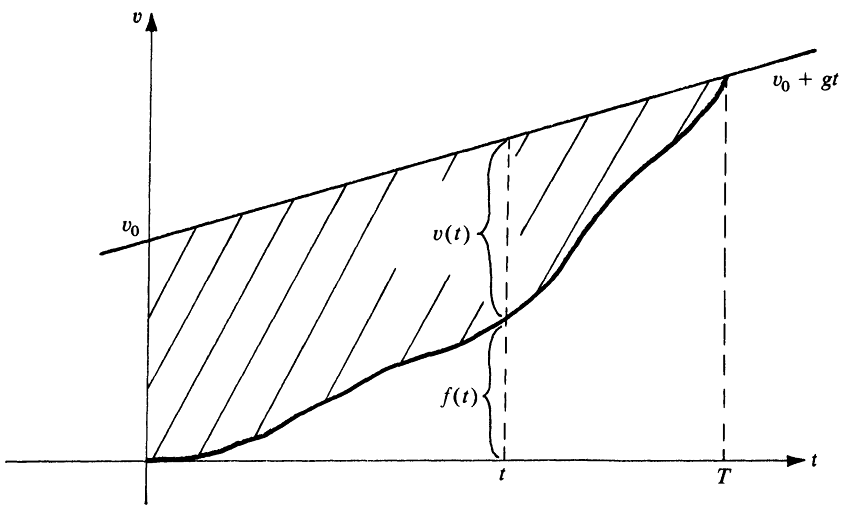



FIG. 1. Meeting a straight line.

Note that the shaded area, between the graph of $f(t)$ and the line $v_0 + gt$, is the integral of $v(t)$ from $t = 0$ to $t = T$, hence is the distance travelled toward the moon. To land, this area must be equal to $x_0$, the starting height.

Therefore, what we must choose is a graph for $f(t)$ which minimizes $T$ and

(1) starts at $f(0) = 0$, since no fuel has been consumed at time $t = 0$,
(2) is nondecreasing, since fuel cannot be unconsumed,
(3) meets the line $v_0 + gt$ at $t = T$ in order to reach $v(T) = 0$, and
(4) encloses, between it and the graph of $v_0 + gt$, an area of $x_0$.

Let us denote by $\underline{M}$ the maximum rate at which fuel can be consumed. In other words, $f'(t) \leq M$ (when $f'(t)$ is defined). Let's see that the graph which minimizes $\underline{T}$ and satisfies conditions (1) through (4) above is $f(t) = 0$ from $t = 0$ to some time $t = T_1$ and then is a straight line of slope $\underline{M}$ from $t = T_1$ to $t = T$, as shown in Fig. 2. This corresponds to waiting as long as possible before firing the rockets, then blasting away at the maximum rate.

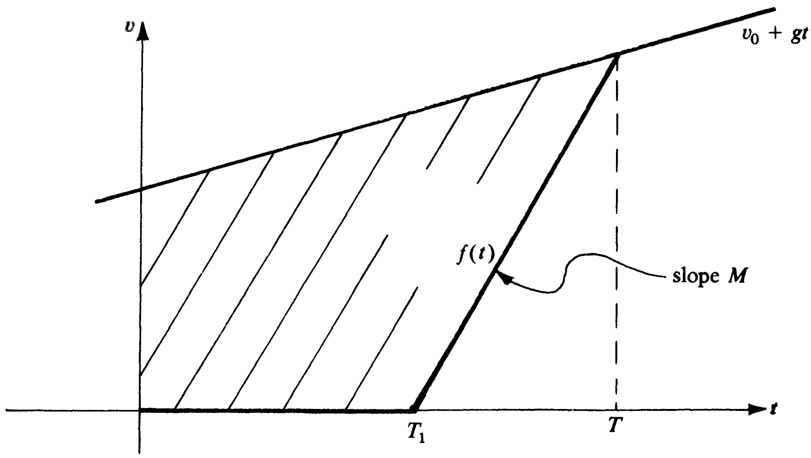



FIG. 2. Holding off, then blasting away.

First note that there is at most one graph of the form described in the previous paragraph which satisfies condition (4), since the area enclosed by such a graph is an increasing function of $T_1$. (If the choice $T_1 = 0$ already encloses an area larger than $x_0$, then we don't have enough fuel to slow down to $v(T) = 0$ before we crash into the moon.) Suppose the graph of a different function $g(t)$ also satisfies conditions (1) through (4) with $g'(t) \leq M$ and meets $v_0 + gt$ at the same time $\underline{T}$, as shown in Fig. 3. (If it meets the line even earlier, we can just continue it along the line to the point where $t = T$.)

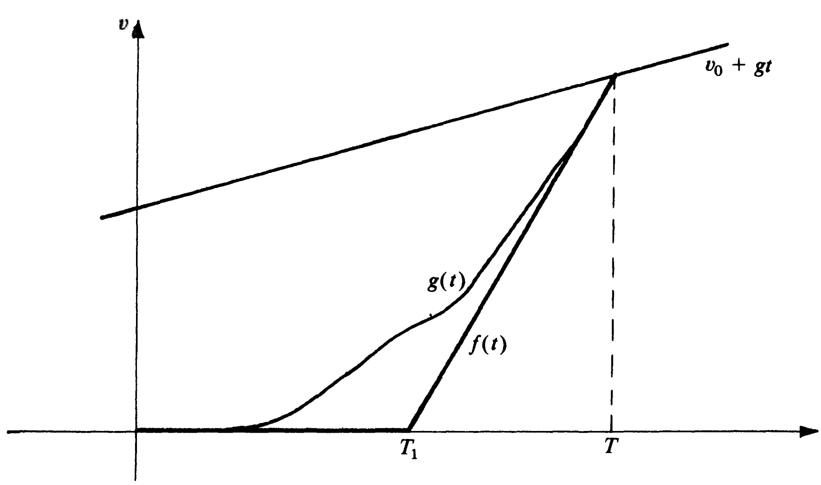



FIG. 3. A different solution?

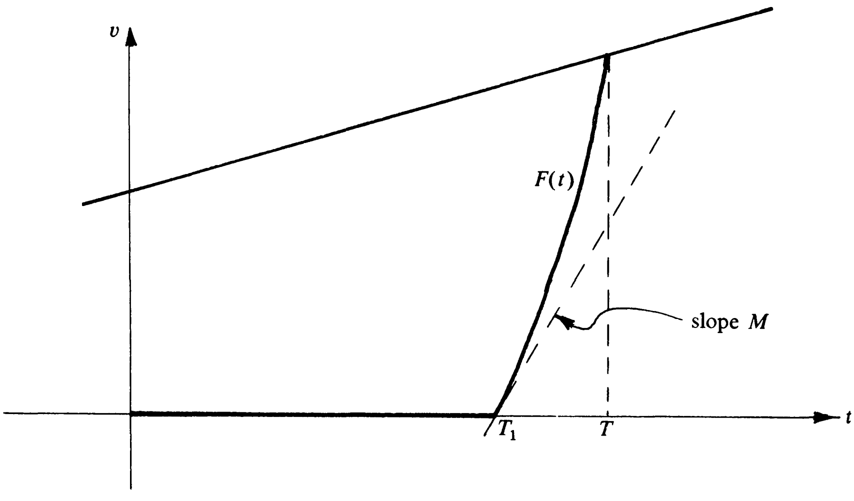



FIG. 4. Maximum firing.

REFERENCES

1. L. Cesari, Optimization Theory and Applications, Springer-Verlag, 1983.
2. R. Holmes, Geometric Functional Analysis and its Applications, Springer-Verlag, 1975.

## A Characterization of a Class of Composition Operators

R. E. LEWKOWICZ
*Department of Mathematics and Statistics, Wright State University, Dayton, Ohio 45435*

A composition operator, or substitution operator, as it is sometimes called, is an operator $T$ on a function space $F(X)$ induced by a transformation $t$ of the underlying space $X$.

$$Tf(x) = f(t(x)).$$

If the function space $F(X)$ is an algebra as well as a vector space and if $T$ preserves the multiplicative properties of $F(X)$ and $t$ is a bijection, then $T$ is an algebra automorphism.

In particular, if $F(K)$ is the space of all real or complex valued functions on a compact Hausdorff space $K$ endowed with pointwise operations and with the topology of pointwise convergence, and $h$ is a homeomorphism of $K$, then the composition operator $H$ defined by $Hf(x) = f(h^{-1}(x))$ is (1) a continuous automorphism of $F(K)$ and (2) preserves the subalgebra of all continuous functions in $F(K)$. The converse is also true:

THEOREM. *If $H$ is a continuous automorphism of $F(K)$ and $H$ preserves the subalgebra of all continuous functions in $F(K)$, then there exists a homeomorphism $h$ of $K$ such that*

$$Hf(x) = f\big(h^{-1}(x)\big)$$

*for all $x$ in $K$.*

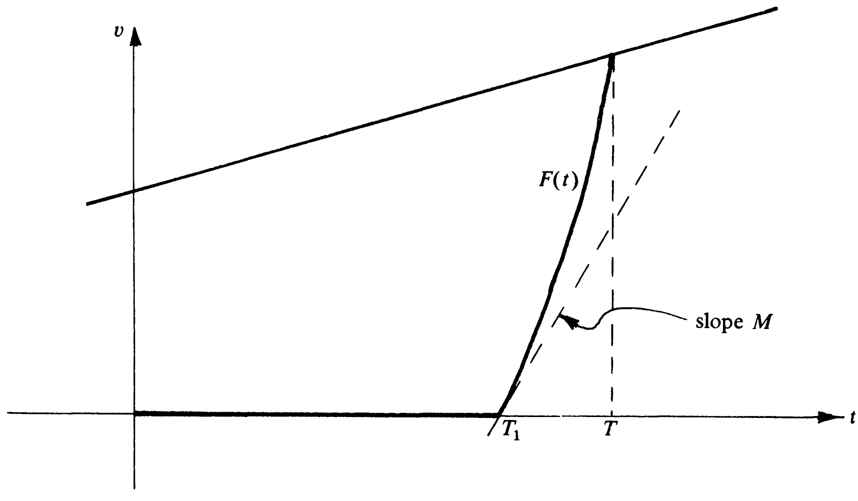



FIG. 4. Maximum firing.

REFERENCES


1. L. Cesari, Optimization Theory and Applications, Springer-Verlag, 1983.
2. R. Holmes, Geometric Functional Analysis and its Applications, Springer-Verlag, 1975.


## A Characterization of a Class of Composition Operators


R. E. LEWKOWICZ

*Department of Mathematics and Statistics, Wright State University, Dayton, Ohio 45435*


A composition operator, or substitution operator, as it is sometimes called, is an operator $T$ on a function space $F(X)$ induced by a transformation $t$ of the underlying space $X$.

$$Tf(x) = f(t(x)).$$

If the function space $F(X)$ is an algebra as well as a vector space and if $T$ preserves the multiplicative properties of $F(X)$ and $t$ is a bijection, then $T$ is an algebra automorphism.

In particular, if $F(K)$ is the space of all real or complex valued functions on a compact Hausdorff space $K$ endowed with pointwise operations and with the topology of pointwise convergence, and $h$ is a homeomorphism of $K$, then the composition operator $H$ defined by $Hf(x) = f(h^{-1}(x))$ is (1) a continuous automorphism of $F(K)$ and (2) preserves the subalgebra of all continuous functions in $F(K)$. The converse is also true:

THEOREM. *If $H$ is a continuous automorphism of $F(K)$ and $H$ preserves the subalgebra of all continuous functions in $F(K)$, then there exists a homeomorphism $h$ of $K$ such that*

$$Hf(x) = f\left(h^{-1}(x)\right)$$

*for all $x$ in $K$.*

*Proof.* If $\chi_E$ is the characteristic function of the subset $E$ of $K$, then by virtue of the multiplicative property of $H$, the function $H\chi_E$ is an idempotent and as such is the characteristic function of a unique subset $\bar{h}(E)$ of $K$. In particular, for each $x \in K$, the image of the characteristic function of the singleton $\{x\}$ is itself the characteristic function of a unique singleton $\{y\}$. Indeed, since $\bar{h}(\{x\}) \neq \varnothing$, there exists an element $y$ in $K$ such that $\bar{h}(\{x\}) = \{y\} \cup P$, where $\{y\} \cap P = \varnothing$. By linearity, invertibility, and the multiplicative property, we find that

$$H^{-1}\chi_{\{y\}} + H^{-1}\chi_P = \chi_{\{x\}}.$$

Hence, if $z \in K$ and $z \neq x$, then

$$H^{-1}\chi_{\{y\}}(z) = 0 \quad \text{and} \quad H^{-1}\chi_P(z) = 0,$$

while if $z = x$, then either

$$H^{-1}\chi_{\{y\}}(z) = 0 \quad \text{and} \quad H^{-1}\chi_P(z) = 1$$

or

$$H^{-1}\chi_{\{y\}}(z) = 1 \quad \text{and} \quad H^{-1}\chi_P(z) = 0.$$

Accordingly, either $H^{-1}\chi_P = \chi_{\{x\}}$ or else $H^{-1}\chi_{\{y\}} = \chi_{\{x\}}$. Since the first of these relations implies that $\chi_{\{y\}} = 0$, the second must hold, implying that $P = \varnothing$.

Define the bijective point function $h: K \to K$ by $h(x) = y$, so that $H\chi_{\{x\}} = \chi_{\{h(x)\}}$. It remains only to show that $Hf = f \circ h^{-1}$ and that $h$ is a homeomorphism. (Since $K$ is compact, it will be sufficient to show that $h^{-1}$ is continuous). The very short and elegant version to be presented here of the remaining part of the proof is due to Eric Nordgren (Private Communication).

For arbitrary $f \in F(K)$ and $x \in K$, we have $f\chi_{\{x\}} = f(x)\chi_{\{x\}}$. Hence

$$\begin{aligned}(Hf)\chi_{\{h(x)\}} &= (Hf)\big(H\chi_{\{x\}}\big) = H\big(f\chi_{\{x\}}\big)\\ &= H\big(f(x)\chi_{\{x\}}\big) = f(x)H\chi_{\{x\}} = f(x)\chi_{\{h(x)\}}.\end{aligned}$$

Evaluation at $h(x)$ now yields $(Hf)(h(x)) = f(x)$ or, equivalently, $Hf = f \circ h^{-1}$. Since $H$ preserves continuity, then $f \circ h^{-1}$ is continuous whenever $f$ is. Since $K$ is compact Hausdorff, this implies that $h^{-1}$ is continuous.

# THE TEACHING OF MATHEMATICS

EDITED BY MELVIN HENRIKSEN AND STAN WAGON

## Coupled Linear Differential Equations with Real Coefficients

MOGENS ESROM LARSEN
*Københavns Universitets Matematiske Institut, Universitetsparken 5, DK-2100 København Ø, Denmark*

BJARNE SLOTH JENSEN
*Copenhagen School of Economics and Business Administration, Copenhagen, Denmark*

**Introduction.** In a first course the case of two coupled linear differential equations tends to fall between two stools. The teacher's unrequited love for eigenvalues drives him into the complex domain, a maze in which he seldom finds the simple, real solutions of the original problem. And even if the complex numbers can be avoided he has difficulty returning through the coordinate transforms. It would seem that if the students had an adequate basis in algebra, everything would be easy. However, on the one hand, it is too much to include all that algebra. On the other hand, that particular subject is not something that can be used now and explained later.

Hence, it is tempting to look for a simple, direct solution, which works in the real domain and only requires straightforward ideas.

**The problem.** We want to analyze an initial value problem: a couple of linear first-order differential equations with constant real coefficients in order to find the real solutions. The system is

$$\dot{x}_1 = ax_1 + bx_2 \tag{1}$$

$$\dot{x}_2 = cx_1 + dx_2 \tag{2}$$

where $a, b, c, d \in \mathbb{R}$, and $x_1, x_2$ are functions with initial values

$$x_1(0) = x_1^0 \tag{3}$$

$$x_2(0) = x_2^0 \tag{4}$$

with $x_1^0, x_2^0 \in \mathbb{R}$. We shall prefer to write it in matrix form. We define vectors

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \qquad \mathbf{x}^0 = \begin{pmatrix} x_1^0 \\ x_2^0 \end{pmatrix}$$

and coefficient matrix

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \tag{5}$$

Then we may replace (1), (2) by (6) and (3), (4) by (7):

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} \tag{6}$$

$$\mathbf{x}(0) = \mathbf{x}^0. \tag{7}$$

**Motivation.** In the traditional search for solutions we argue along the following lines: If $\mathbf{A}$ should happen to be a diagonal matrix, i.e., $b = c = 0$, then the system consists of two independent equations, namely

$$\dot{x}_1 = ax_1, \quad \dot{x}_2 = dx_2,$$

with independent inital values

$$x_1(0) = x_1^0, \quad x_2(0) = x_2^0.$$

If $\mathbf{A}$ is not diagonal, we look for a coordinate transformation

$$\mathbf{x} = \mathbf{S}\mathbf{y}$$

which changes the equation to

$$\dot{\mathbf{y}} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}\mathbf{y}. \tag{8}$$

If the new matrix

$$\mathbf{B} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}$$

happens to be diagonal, then we are through. Unfortunately, we might need to extend the problem into the complex domain in order to obtain this diagonalization, and even so, as the matrix

$$\begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}$$

shows, not all matrices can be diagonalized. In spite of the large amount of algebra employed we have hardly succeeded in finding the real solutions.

**Alternative analysis.** The idea to be explained below is to argue slightly differently: If $\mathbf{A}^2$ should happen to be diagonal, then the system is easy to solve, even as an initial value problem. If $\mathbf{A}^2$ is not diagonal, then we are able to transform the problem, such that the new one has a coefficient matrix with diagonal square. As a matter of fact, neither of the above features needs complex numbers, and further, there are no exceptions to the procedure or even to the formulas for the solutions of the inital value problem.

**The solution with trace zero.** If $\mathbf{A}$ is not already a diagonal matrix, then $\mathbf{A}^2$ is diagonal if and only if the trace $\mathbf{A}$ is zero. As we shall see, we can always transform the problem to the case where the trace of the new coefficient matrix is zero, even when $\mathbf{A}$ is diagonal. Hence we shall restrict our analysis to the case of trace zero.

THEOREM 1. *If* $\mathbf{A}$ *has trace zero, then* $-\mathbf{A}^2$ *is the determinant of* $\mathbf{A}$ *times the unit matrix, i.e.,* $\mathbf{A}^2 = \Delta\mathbf{E}$ *with* $\Delta = a^2 + bc$.

THEOREM 2. *If* $\mathbf{A}^2$ *is diagonal, then either* $\mathbf{A}$ *is diagonal or the trace of* $\mathbf{A}$ *is zero.*

*Proof.* Elementary. ∎

Under this assumption we shall analyse the solution of (6) and (7). Let $\mathbf{x}$ be a solution of (6)–(7). Then by Theorem 1

$$\ddot{\mathbf{x}} = \Delta\mathbf{E}\mathbf{x}. \tag{9}$$

Let $\delta$ be the solution of the initial value problem

$$\ddot{\delta} = \Delta\delta; \tag{10}$$

$$\delta(0) = 0, \qquad \dot{\delta}(0) = 1. \tag{11}$$

Note that $\dot{\delta}$ solves (10), but not (11). Then $\dot{\delta}$ is not proportional to $\delta$ and hence the couple $(\delta, \dot{\delta})$ constitutes a basis for the solution of (10). Because $\mathbf{x}$ solves (9), it must take the form

$$\mathbf{x} = \dot{\delta}\mathbf{v} + \delta\mathbf{w} \tag{12}$$

where $\mathbf{v}$ and $\mathbf{w}$ are vectors in $\mathbb{R}^2$. As $\mathbf{x}$ satisfies (6), we have

$$\dot{\mathbf{x}} = \ddot{\delta}\mathbf{v} + \dot{\delta}\mathbf{w} = \mathbf{A}\mathbf{x} = \dot{\delta}\mathbf{A}\mathbf{v} + \delta\mathbf{A}\mathbf{w}.$$

Using (10) we get the equation

$$\dot{\delta}\mathbf{w} + \delta\Delta\mathbf{v} = \dot{\delta}\mathbf{A}\mathbf{v} + \delta\mathbf{A}\mathbf{w}.$$

At $t = 0$ we have, because of (11),

$$\mathbf{w} = \mathbf{A}\mathbf{v}. \tag{13}$$

Substitution of (13) in (12) yields

$$\mathbf{x} = (\dot{\delta}\mathbf{E} + \delta\mathbf{A})\mathbf{v}.$$

As $\mathbf{x}$ satisfies (7), we have

$$\mathbf{x}^0 = \mathbf{x}(0) = (1\mathbf{E} + 0\mathbf{A})\mathbf{v} = \mathbf{v}. \tag{14}$$

Hence the solution of (6) and (7) is of the form

$$\mathbf{x} = (\dot{\delta}\mathbf{E} + \delta\mathbf{A})\mathbf{x}^0. \tag{15}$$

This ends the analysis.

Now we can substitute (15) in (6) and (7) for verification. In the latter case we get (14), and in the former (using (10))

$$\dot{\mathbf{x}} = (\ddot{\delta}\mathbf{E} + \dot{\delta}\mathbf{A})\mathbf{x}^0 = (\delta\Delta\mathbf{E} + \dot{\delta}\mathbf{A})\mathbf{x}^0,$$

while using Theorem 1 yields

$$\mathbf{A}\mathbf{x} = \mathbf{A}(\dot{\delta}\mathbf{E} + \delta\mathbf{A})\mathbf{x}^0 = (\dot{\delta}\mathbf{A} + \delta\mathbf{A}^2)\mathbf{x}^0 = (\dot{\delta}\mathbf{A} + \delta\Delta\mathbf{E})\mathbf{x}^0.$$

**The general case.** Without any assumptions about $\mathbf{A}$ we shall transform the equations (6) and (7) to the case of zero trace. This proves much easier than the transformation (8). Let $\Theta \in \mathbb{R}$ be a constant, and $\theta$ the solution of the initial value problem

$$\dot{\theta} = \Theta\theta; \qquad \theta(0) = 1 \tag{16}$$

(i.e., the exponential function $\theta(t) = e^{\Theta t}$).

We consider the coordinate transformation

$$\mathbf{x} = \theta\boldsymbol{\xi}.$$

Now (6) and (7) for $\mathbf{x}$ imply certain equations for $\boldsymbol{\xi}$. (7) is simple:

$$\mathbf{x}^0 = \mathbf{x}(0) = \theta(0)\boldsymbol{\xi}(0) = \boldsymbol{\xi}(0)$$

by (16). (6) is nicer:

$$\dot{\mathbf{x}} = \theta\dot{\boldsymbol{\xi}} + \dot{\theta}\boldsymbol{\xi} = \mathbf{A}\mathbf{x} = \theta\mathbf{A}\boldsymbol{\xi}.$$

Using (16) we get

$$\theta\dot{\boldsymbol{\xi}} + \Theta\theta\boldsymbol{\xi} = \theta\mathbf{A}\boldsymbol{\xi}.$$

Because $\theta \neq 0$, we can divide by it, hence,

$$\dot{\boldsymbol{\xi}} = (\mathbf{A} - \Theta\mathbf{E})\boldsymbol{\xi}.$$

Now, if we choose $\Theta$ correctly, the new matrix will have trace zero. We define $\Theta$ as

$$\Theta = \frac{a+d}{2},$$

half of the trace of $\mathbf{A}$. Then the system gets the matrix

$$\mathbf{B} = (\mathbf{A} - \Theta\mathbf{E}) = \begin{pmatrix} \dfrac{a-d}{2} & b \\ c & \dfrac{d-a}{2} \end{pmatrix}$$

so $\boldsymbol{\xi}$ solves the initial-value problem:

$$\dot{\boldsymbol{\xi}} = \mathbf{B}\boldsymbol{\xi}; \qquad \boldsymbol{\xi}(0) = \mathbf{x}^0 \tag{17}$$

of the type of trace zero. (17) is then solved by (15), where $\delta$ solves (10), (11) with

$$\Delta = \left(\frac{a-d}{2}\right)^2 + bc.$$

**Conclusion.** We can write down the solution of (6) and (7) explicitly. Let the half-trace $\Theta$ and the discriminant $\Delta$ of the matrix $\mathbf{A}$ be defined as

$$\Theta = \frac{a+d}{2}, \tag{18}$$

$$\Delta = \left(\frac{a-d}{2}\right)^2 + bc. \tag{19}$$

Let $\theta$ be the solution of the initial-value problem

$$\dot{\theta} = \Theta\theta; \qquad \theta(0) = 1.$$

Let $\delta$ be the solution of the initial-value problem

$$\ddot{\delta} = \Delta\delta; \qquad \delta(0) = 0; \qquad \dot{\delta}(0) = 1.$$

Then the solution of (6) and (7) is:

$$\mathbf{x} = \theta\big(\dot{\delta}\mathbf{E} + \delta(\mathbf{A} - \Theta\mathbf{E})\big)\mathbf{x}^0. \tag{20}$$

In coordinates this becomes

$$x_1 = \theta\left(x_1^0\dot{\delta} + \left(\frac{a-d}{2}x_1^0 + bx_2^0\right)\delta\right),$$

$$x_2 = \theta\left(x_2^0\dot{\delta} + \left(\frac{d-a}{2}x_2^0 + cx_1^0\right)\delta\right).$$

The functions $\theta$ and $\delta$ can be explicitly written down. They are

$$\theta(t) = e^{\Theta t} = e^{\frac{a+d}{2}t};$$

$$\delta(t) = \begin{cases} \dfrac{1}{\sqrt{\Delta}}\sinh(\sqrt{\Delta}\,t) & \text{for } \Delta > 0, \\ t & \text{for } \Delta = 0, \\ \dfrac{1}{\sqrt{-\Delta}}\sin(\sqrt{-\Delta}\,t) & \text{for } \Delta < 0. \end{cases}$$

**Afterthought.** From a higher point of view, the methods applied here are examples of more sophisticated analytic methods in algebraic disguise, to be compared with the standard sophisticated algebra. If Sophus Lie could have asked Jean B. J. Fourier to solve the equations, he would have done so as follows:

The system (1)–(2) should be transformed into one equation of second order, i.e.,

$$\ddot{x} - (a + d)\dot{x} + (ad - bc)x = 0.$$

Fourier, of course, would have transformed the operator to a polynomial,

$$\xi^2 - (a + d)\xi + (ad - bc);$$

then he would have translated this by the distance $\Theta$ (from (18)), say

$$\eta = \xi - \Theta, \tag{21}$$

and, hence, obtained

$$\eta^2 - \Theta^2 + (ad - bc) = \eta^2 - \Delta$$

with $\Delta$ defined by (19).

By the inverse Fourier transformation, $\eta$ is transformed into $y$, satisfying

$$\ddot{y} = \Delta y$$

and related to $x$ by the transform of (21), i.e.,

$$y = x \cdot e^{-\Theta t}.$$

Further, he would have formulated the results of his efforts in the form of (20). For then Sophus Lie could have extracted the matrix

$$\mathbf{C}(t) = \theta(t)\big(\dot{\delta}(t)\mathbf{E} + \delta(t)(\mathbf{A} - \Theta\mathbf{E})\big),$$

which is a handy representation of the *Lie group* of the flow of solutions of (1)–(4).

Hence $\mathbf{C}(t)$ must satisfy the relation

$$\mathbf{C}(t + s) = \mathbf{C}(t)\mathbf{C}(s).$$

We know this relation from the theory of Lie groups, but we shall leave verification by the elementary trigonometric formulas for addition and their analogues as an exercise.

REFERENCES

1. Witold Hurewitz, Lectures on Ordinary Differential Equations, Wiley, New York, 1958, pp. 34–44.
2. E. L. Ince, Ordinary Differential Equations, Dover, New York, 1944, pp. 133–155.

# Irrationals and the Fundamental Theorem of Arithmetic

DAVID J. SPROWS
*Department of Mathematical Sciences, Villanova University, Villanova, PA 19085*

Recent issues of the MONTHLY contain articles [**1**, **2**] that discuss approaches to the irrationality of the square roots of nonsquare natural numbers using arguments designed to avoid any mention of unique factorization. The argument in [**2**] is incomplete but, as the article points out, the important objective when introducing a topic such as the existence of irrationals is to present arguments that the students find convincing and many students have difficulty with reasoning involving the unique factorization aspect of the Fundamental Theorem of Arithmetic. The following discussion is designed to give students some appreciation of the significance of the property of unique factorization by considering the "irrationality" of square roots in systems where this property does not hold. Ideally it should come immediately after the presentation of a plausibility argument such as that in [**2**] showing the students that square roots of primes are irrational.

Given a subset $A$ of the natural numbers that is closed under multiplication, we say an element $c$ of $A$ is *$A$-composite* provided $c = ab$ where $a$ and $b$ are elements in $A$ with $a \neq 1$ and $b \neq 1$. An element $c$ in $A$ is called *$A$-prime* provided $c \neq 1$ and $c$ is not $A$-composite. For example, if $A$ is the set of all multiples of 4, then 16, 48 and 80 are $A$-composites and 4, 40 and 100 are $A$-primes.

We say a real number $x$ is *$A$-rational* provided $x$ can be expressed in the form $a/b$ where $a$ and $b$ are elements of $A$. Note that each element $c$ in $A$ is $A$-rational since $c$ can be expressed as $c^2/c$. This is analogous to the fact that every natural number is rational. However, there are properties of the natural numbers that do not extend to any set $A$. For example, let $A$ be the set of all multiples of 4 and let $c = 100$. Since $c$ is an $A$-prime, by analogy with the natural numbers its square root should not be $A$-rational, but $\sqrt{c} = a/b$ where $a = 40$ and $b = 4$. Such a situation is possible since in the set $A$ the number $a^2$ can be factored into $A$-primes as either $a \cdot a$ or $c \cdot b \cdot b$.

There is nothing special about the multiples of 4 in the above example. A useful exercise is to have each student pick a subset of the natural numbers that is closed under multiplication and investigate the "irrationality" of roots in this system. This serves to illustrate the rarity of the unique factorization property and the crucial role played by this property in the study of algebraic irrationals.

REFERENCES

1. Y. Sagher, What Pythagoras could have done, this MONTHLY, 95 (1988) 117.
2. W. C. Waterhouse, Why square roots are irrational, this MONTHLY, 93 (1986) 213–214.

Prove that two of these three fields $R$, $S$, $T$ are equal and that the remaining one is a proper subfield of the other two.

E3346. *Proposed by Dean S. Clark, University of Rhode Island, Kingston.*

Call a subset $T$ of $\mathbb{Z}$ mod $n$ *sum-free* if the sum of two distinct elements of $T$ is not in $T$. Let $s(n)$ be the maximum cardinality of a sum-free subset of $\mathbb{Z}$ mod $n$.

(a) Prove that $s(n) = n/2$ if $n$ is even.

(b) If $n$ is odd, prove that $s(n) \geqslant \lfloor n/3 \rfloor + 1$.

(c)* If $n$ is odd, is $s(n) = \lfloor n/3 \rfloor + 1$?

E3347. *Proposed by Vlastimil Dlab, Carleton University, Ottawa, Canada.*

Given an arbitrary natural number $N$, show that every rational number between 0 and 1 can be written as a finite product of numbers of the form

$$R_n = \frac{n(n+3)}{(n+1)(n+2)}, \qquad n \geqslant N.$$

E3348. *Proposed by Alfred Witkowski, University of Wrocław, Poland.*

(a) Suppose $k$ is a given nonnegative integer. If $n > k$ and

$$P_n(x) = \prod_{j=0}^{n} (x - j),$$

let $x_{k,n}$ be the zero of $P_n'$ in the interval $(k, k+1)$. Find $\lim_{n\to\infty} x_{k,n}$ and estimate the speed of convergence.

(b) If $n > 0$ and

$$Q_n(x) = \prod_{j=0}^{n} (x - j^2),$$

let $t_n$ be the zero of $Q_n'$ in the interval $(0,1)$. Find $\lim_{n\to\infty} t_n$ and estimate the speed of convergence.

## SOLUTIONS OF ELEMENTARY PROBLEMS

### Lobachevskian Altitudes

E 3185 [1987, 71]. *Proposed by Raymond E. Spaulding, Radford University, Radford, VA.*

Let $P$ be a point in the interior of an equilateral triangle, and let $S$ be the sum of the perpendicular distances to the three sides of the triangle from $P$. In Euclidean geometry, the sum $S$ always equals the altitude of the triangle. In Lobachevskian geometry, prove that $S$ is less than any altitude. In addition, find the position of $P$ which would give a minimum value for $S$.

*Solution by the proposer.*

LEMMA 1. *If $ABCD$ is a quadrilateral with $A, B, C$ having measure $\pi/2$, and $P$ is any point in the interior with perpendiculars to $AD$ and $BC$ at $A'$ and $B'$, then $AB < PA' + PB' < DC$.*

*Proof.* Drop a perpendicular from $A'$ to $BC$. The foot $B''$ will be between $B$ and $B'$, since otherwise a triangle with angle sum exceeding $\pi$ will be formed. Exercise I.9 of *Basic Concepts of Geometry* by Prenowitz and Jordan (p. 86) states that $AB < A'B'' < DC$. By using triangles, it is easy to show that $A'B'' < PA' + PB' < DC$, and the result follows. □

Given a triangle $ABC$ and a point $P$ on its perimeter or interior, let $d(P)$ denote the sum of the perpendicular distances from $P$ to the sides of the triangle.

LEMMA 2. *In a triangle $ABC$ with $\angle A < \pi/2$ and $AB = AC$, let $P_0$ be the foot of the perpendicular from $A$. If $P_1$ is any point between $B$ and $C$ other than $P_0$, then $d(P_0) < d(P_1) < h$, where $h$ is the distance from $B$ to side $AC$.*

*Proof.* Reflect $A$ through $BC$ to $A'$, obtaining $\triangle A'BC$ congruent to $\triangle ABC$. For $i \in \{0,1\}$, let $B_i$ and $C_i$ be the feet of the perpendiculars from $P_i$ to $AC$ and $AB$, respectively, and extend $B_iP_i$ to meet $BA'$ at $D_i$. Let $B_2, C_2$ be the feet of the perpendiculars from $B$ to $AC$ and $C$ to $BA'$, respectively. Then $\triangle P_0B_0C = \triangle P_0D_0B$ by ASA, $\triangle P_1C_1B = \triangle P_1D_1B$ by AAS, and $\triangle BB_2C = \triangle CC_2B$ by AAS. Thus $CC_2 = BB_2$. By Lemma 1, $D_0B_0 < D_1P_1 + P_1B_1 < CC_2$, which means that $P_0C_0 + P_0B_0 < P_1C_1 + P_1B_1 < BB_2$. □

Now let $ABC$ be an equilateral triangle with altitude $AA_0$, and let $P_1$ be a point inside $ABC$ on the $C$ side of $AA_0$. Suppose the perpendicular from $P_1$ to $AA_0$ meets $AA_0$ at $P_0$ and $AB$ at $P_2$. For $i \in \{0,1,2\}$, let $A_i, B_i, C_i$ be the feet of the perpendiculars from $P_i$ to $BC$, $AC$, $AB$, respectively. By Lemma 2, $P_0C_0 + P_0B_0 < P_1C_1 + P_1B_1 < P_2B_2$. By the exercise previously quoted, $P_0A_0 < P_1A_1 < P_2A_2$. Summing these yields $P_0C_0 + P_0B_0 + P_0A_0 < P_1C_1 + P_1B_1 + P_1A_1 < P_2B_2 + P_2A_2$. By Lemma 2, $P_2B_2 + P_2A_2$ is less than any altitude of equilateral triangle $ABC$. This proves that $S = d(P)$ is less than any altitude for any $P$ in the triangle, and also that $S$ is minimized uniquely by the intersection of the altitudes from the three vertices.

No other solutions were received.

**3-Player Blackjack**

E 3186 [1987, 71]. *Proposed by I. A. Sakmar, University of Western Ontario.*

On a current TV game show, three contestants for the final showcase prizes compete by spinning a wheel, which is divided into twenty equal sections numbered in steps of 5 from 5 to 100. The aim is to get as close to 100 as possible without exceeding a sum of 100 in two spins. Each contestant completes one or both spins before the next contestant's turn, and if either the 1st or 2nd contestant exceeds 100, that contestant is out of contention. In case of a tie, the first contestant to attain that score is chosen.

*Solution by the proposer.*

LEMMA 1. *If $ABCD$ is a quadrilateral with $A, B, C$ having measure $\pi/2$, and $P$ is any point in the interior with perpendiculars to $AD$ and $BC$ at $A'$ and $B'$, then $AB < PA' + PB' < DC$.*

*Proof.* Drop a perpendicular from $A'$ to $BC$. The foot $B''$ will be between $B$ and $B'$, since otherwise a triangle with angle sum exceeding $\pi$ will be formed. Exercise I.9 of *Basic Concepts of Geometry* by Prenowitz and Jordan (p. 86) states that $AB < A'B'' < DC$. By using triangles, it is easy to show that $A'B'' < PA' + PB' < DC$, and the result follows. □

Given a triangle $ABC$ and a point $P$ on its perimeter or interior, let $d(P)$ denote the sum of the perpendicular distances from $P$ to the sides of the triangle.

LEMMA 2. *In a triangle $ABC$ with $\angle A < \pi/2$ and $AB = AC$, let $P_0$ be the foot of the perpendicular from $A$. If $P_1$ is any point between $B$ and $C$ other than $P_0$, then $d(P_0) < d(P_1) < h$, where $h$ is the distance from $B$ to side $AC$.*

*Proof.* Reflect $A$ through $BC$ to $A'$, obtaining $\triangle A'BC$ congruent to $\triangle ABC$. For $i \in \{0,1\}$, let $B_i$ and $C_i$ be the feet of the perpendiculars from $P_i$ to $AC$ and $AB$, respectively, and extend $B_iP_i$ to meet $BA'$ at $D_i$. Let $B_2, C_2$ be the feet of the perpendiculars from $B$ to $AC$ and $C$ to $BA'$, respectively. Then $\triangle P_0B_0C = \triangle P_0D_0B$ by ASA, $\triangle P_1C_1B = \triangle P_1D_1B$ by AAS, and $\triangle BB_2C = \triangle CC_2B$ by AAS. Thus $CC_2 = BB_2$. By Lemma 1, $D_0B_0 < D_1P_1 + P_1B_1 < CC_2$, which means that $P_0C_0 + P_0B_0 < P_1C_1 + P_1B_1 < BB_2$. □

Now let $ABC$ be an equilateral triangle with altitude $AA_0$, and let $P_1$ be a point inside $ABC$ on the $C$ side of $AA_0$. Suppose the perpendicular from $P_1$ to $AA_0$ meets $AA_0$ at $P_0$ and $AB$ at $P_2$. For $i \in \{0,1,2\}$, let $A_i, B_i, C_i$ be the feet of the perpendiculars from $P_i$ to $BC$, $AC$, $AB$, respectively. By Lemma 2, $P_0C_0 + P_0B_0 < P_1C_1 + P_1B_1 < P_2B_2$. By the exercise previously quoted, $P_0A_0 < P_1A_1 < P_2A_2$. Summing these yields $P_0C_0 + P_0B_0 + P_0A_0 < P_1C_1 + P_1B_1 + P_1A_1 < P_2B_2 + P_2A_2$. By Lemma 2, $P_2B_2 + P_2A_2$ is less than any altitude of equilateral triangle $ABC$. This proves that $S = d(P)$ is less than any altitude for any $P$ in the triangle, and also that $S$ is minimized uniquely by the intersection of the altitudes from the three vertices.

No other solutions were received.

**3-Player Blackjack**

E 3186 [1987, 71]. *Proposed by I. A. Sakmar, University of Western Ontario.*

On a current TV game show, three contestants for the final showcase prizes compete by spinning a wheel, which is divided into twenty equal sections numbered in steps of 5 from 5 to 100. The aim is to get as close to 100 as possible without exceeding a sum of 100 in two spins. Each contestant completes one or both spins before the next contestant's turn, and if either the 1st or 2nd contestant exceeds 100, that contestant is out of contention. In case of a tie, the first contestant to attain that score is chosen.

Spinning again increases the probability of winning if and only if $x \leqslant 13$, so the first player stops on the first spin if she spins at least 14. Her probability of winning is

$$\sum_{i=1}^{20} \frac{1}{20} \max\{f_2(i), F_2(i)\} \cong .341771.$$

*Editorial comment.* One can consider the more general game with $n$ successive players. For $n = 2$, the optimal stopping value for the first player is 11, as computed above, and the win probability is $\Sigma_{i=1}^{20}(1/20)\max\{f_1(i), F_1(i)\} \cong .482437$. As $n$ increases, the optimal stopping value for the first player increases to 20, and her win probability decreases to $(1/20) + (19/20)(1/20) = .0975$.

Most solvers assumed without comment that a player would not stop on the first spin if and only if she did not exceed all previous players that did not exceed 20. Because the correct stopping value for the first player is above that of the second player, this does not affect the computation of the stopping value, but it gives too small a win probability for the first player. Such an assumption is considered a partial solution. It is also possible to argue more globally that the stopping values form a decreasing sequence in order to justify the use of $f_2(x) = x^4/400$ in the range of interest. In order to win, the first player needs a value high enough to give a reasonable chance of beating both of the others. In general, it will have to be higher than that required to beat only the last player.

David Bloom commented on the replacement of 20 by $m$ and let $m$ approach infinity, obtaining polynomial equations for the optimal stopping values in the case $n = 3$. This analysis is equivalent to that of the continuous game proposed by Peter Griffin. Each contestant draws a random number uniformly in [0, 1], with the option to draw again, losing if the sum of the two draws exceeds 1. For $n = 3$, reasoning as for the discrete game yields $f_1(y) = y^2$ and $F_1(y) = \int_y^1 t^2\,dt = (1 - y^3)/3$, yielding a critical value for $B$ of $y_0 \cong .532089$ as the solution to $y^3 + 3y^2 = 1$. For $A$, we then have $f_2(x) = x^4$ if $x > y_0$, and $f_2(x) = x^2(y_0^2 + x^2)/2$ if $x < y_0$. Comparing this with $F_2(x) = \int_{x_0}^1 f_2(t)\,dt$ yields a critical value for $A$ of $x_0 = .648655$ as the solution to $x^5 + 5x^4 = 1$. The probability that $A$ wins is $\int_0^{x_0} F_2(t)\,dt + \int_{x_0}^1 f_2(t)\,dt = (1 + x_0 - x_0^5)/5 - x_0^6/30 + y_0^6/24 \cong .305227$. The probability that the first player wins the 2-person game is .453802. In each case this is less than the probability when $m = 20$, showing the effect of the rule about ties.

Also solved by M. Orlowski and M. Pachter (South Africa) and John Lawrence. Partially solved by D. Callan, T. E. Elsner, P. Griffin, M. O. Marchand, J. H. Steelman, O. P. Lossers (The Netherlands), and the proposer. Two incorrect solutions were received.

### A Sufficient Condition for Similarity

E 3193 [1987, 181]. *Proposed by Andrew Lenard, Indiana University.*

Let $\theta$ be an (undirected) acute angle. Show that if a one-to-one mapping $T$ of the Euclidean plane $E$ onto itself has the property that whenever points $P$ and $Q$ subtend the angle $\theta$ at the point $R$ then also the points $T(P)$ and $T(Q)$ subtend the angle $\theta$ at the point $T(R)$, then $T$ is a similarity transformation of $E$.

*Composite solution by L. E. Mattics, University of South Alabama, the proposer, and the editors.*

Let $X'$ denote the image under $T$ of any point or point set $X$. We prove first that $T$ preserves betweenness. Suppose $A, B, C$ are collinear with $B$ between $A$ and $C$. Along the perpendicular bisector of $AC$, locate symmetric points $M$ and $N$ such that the four angles $MAC$, $MCA$, $NAC$, and $NCA$ all equal $\theta$. This is possible because $\theta$ is acute. Note that $AMCN$ is a rhombus, and that the four angles $MAB$, $MCB$, $NAB$, and $NCB$ also equal $\theta$. By the hypothesized property of $T$, $A'M'C'N'$ must also be a rhombus, and $B'$ must lie on the segment $A'C'$.

Next we show that $T$ takes lines to lines and preserves parallelism. If $P$ lies off line $L$, choose $Q$ and $R$ on $L$ so that angle $PQR$ measures $\theta$. The angle $P'Q'R'$ must also be $\theta$ and $P'$ does not lie on the line through $Q'$ and $R'$. Since $T$ is one-to-one, each point on this line must therefore be the image of a point on $L$, and so $L'$ is the entire line through $Q'$ and $R'$. Also, if $L_1$ and $L_2$ are parallel, then $L_1'$ and $L_2'$ are disjoint and therefore parallel. In particular, note that the image of a parallelogram is a parallelogram.

Next we show that $T$ scales distances uniformly on any given line, i.e., given a line $L$, there is a number $a(L)$ such that $|A'B'| = a(L)|AB|$ for each $A, B \in L$. First consider segments of a fixed length, say 1. Let $A, B, C, D$ be points on $L$ and $R, S$ be points on a line parallel to $L$ such that $|AB| = |CD| = |RS| = 1$. Then $ABRS$ and $CDRS$ are parallelograms (provided the points are named in the correct order). Therefore, $A'B'R'S'$ and $C'D'R'S'$ are also parallelograms, and $|A'B'| = |R'S'| = |C'D'|$. Thus $|A'B'|/|AB|$ is independent of the positions of $A$ and $B$, but perhaps depends on $|AB|$. Let $a = |A'B'|/|AB|$ for $|AB| = 1$. To show $|P'Q'| = a|PQ|$ for arbitrary $P, Q \in L$, let $n$ be an integer. If $|PQ| = n$, divide $PQ$ into unit segments. If $|PQ| = 1/n$, concatenate $n$ copies of $PQ$. If $|PQ|$ is rational, combine these operations. If $|PQ|$ is irrational, trap it between two sequences of rational lengths approaching it from above and below, and use the betweenness property established above.

Finally, we show that $a(L)$ is independent of $L$. By the parallelogram construction above, $a(L_1) = a(L_2)$ if $L_1$ and $L_2$ are parallel. Otherwise, let $A = L_1 \cap L_2$. Choose $B$ on $L_1$ and $C$ on $L_2$ so that $|AB| = 1$ and angle $ABC = \theta$. Also choose $D$ on $L_2$ and $E$ on $L_1$ so that $|AD| = 1$ and angle $ADE = \theta$. Note that $|AC| = |AE|$; let $d$ be their common value. Triangles $A'B'C'$ and $A'D'E'$ are similar, because they have a common angle at $A'$, and angle $A'B'C'$ = angle $A'D'E' = \theta$. Hence $|A'E'|/|A'C'| = |A'D'|/|A'B'|$. Substituting for these values in terms of $d$ and the scaling constants for $L_1$ and $L_2$ yields $d \cdot a(L_1)/\{d \cdot a(L_2)\} = a(L_2)/a(L_1)$. Since the $a$'s are positive, this yields $a(L_1) = a(L_2)$.

No other solutions were received.

### The $n$th Mean Value Theorem

E 3214 [1987, 548]. *Proposed by Guo Qiang Zhang, University of Cambridge, England.*

Let $f$ be a real function with $n+1$ derivatives on $[a, b]$. Suppose $f^{(i)}(a) = f^{(i)}(b) = 0$ for $i = 0, 1, \ldots, n$. Prove that there is a number $\xi$ in $(a, b)$ such that $f^{(n+1)}(\xi) = f(\xi)$.

*Composite solution by L. E. Mattics, University of South Alabama, the proposer, and the editors.*

Let $X'$ denote the image under $T$ of any point or point set $X$. We prove first that $T$ preserves betweenness. Suppose $A, B, C$ are collinear with $B$ between $A$ and $C$. Along the perpendicular bisector of $AC$, locate symmetric points $M$ and $N$ such that the four angles $MAC$, $MCA$, $NAC$, and $NCA$ all equal $\theta$. This is possible because $\theta$ is acute. Note that $AMCN$ is a rhombus, and that the four angles $MAB$, $MCB$, $NAB$, and $NCB$ also equal $\theta$. By the hypothesized property of $T$, $A'M'C'N'$ must also be a rhombus, and $B'$ must lie on the segment $A'C'$.

Next we show that $T$ takes lines to lines and preserves parallelism. If $P$ lies off line $L$, choose $Q$ and $R$ on $L$ so that angle $PQR$ measures $\theta$. The angle $P'Q'R'$ must also be $\theta$ and $P'$ does not lie on the line through $Q'$ and $R'$. Since $T$ is one-to-one, each point on this line must therefore be the image of a point on $L$, and so $L'$ is the entire line through $Q'$ and $R'$. Also, if $L_1$ and $L_2$ are parallel, then $L_1'$ and $L_2'$ are disjoint and therefore parallel. In particular, note that the image of a parallelogram is a parallelogram.

Next we show that $T$ scales distances uniformly on any given line, i.e., given a line $L$, there is a number $a(L)$ such that $|A'B'| = a(L)|AB|$ for each $A, B \in L$. First consider segments of a fixed length, say 1. Let $A, B, C, D$ be points on $L$ and $R, S$ be points on a line parallel to $L$ such that $|AB| = |CD| = |RS| = 1$. Then $ABRS$ and $CDRS$ are parallelograms (provided the points are named in the correct order). Therefore, $A'B'R'S'$ and $C'D'R'S'$ are also parallelograms, and $|A'B'| = |R'S'| = |C'D'|$. Thus $|A'B'|/|AB|$ is independent of the positions of $A$ and $B$, but perhaps depends on $|AB|$. Let $a = |A'B'|/|AB|$ for $|AB| = 1$. To show $|P'Q'| = a|PQ|$ for arbitrary $P, Q \in L$, let $n$ be an integer. If $|PQ| = n$, divide $PQ$ into unit segments. If $|PQ| = 1/n$, concatenate $n$ copies of $PQ$. If $|PQ|$ is rational, combine these operations. If $|PQ|$ is irrational, trap it between two sequences of rational lengths approaching it from above and below, and use the betweenness property established above.

Finally, we show that $a(L)$ is independent of $L$. By the parallelogram construction above, $a(L_1) = a(L_2)$ if $L_1$ and $L_2$ are parallel. Otherwise, let $A = L_1 \cap L_2$. Choose $B$ on $L_1$ and $C$ on $L_2$ so that $|AB| = 1$ and angle $ABC = \theta$. Also choose $D$ on $L_2$ and $E$ on $L_1$ so that $|AD| = 1$ and angle $ADE = \theta$. Note that $|AC| = |AE|$; let $d$ be their common value. Triangles $A'B'C'$ and $A'D'E'$ are similar, because they have a common angle at $A'$, and angle $A'B'C' =$ angle $A'D'E' = \theta$. Hence $|A'E'|/|A'C'| = |A'D'|/|A'B'|$. Substituting for these values in terms of $d$ and the scaling constants for $L_1$ and $L_2$ yields $d \cdot a(L_1)/\{d \cdot a(L_2)\} = a(L_2)/a(L_1)$. Since the $a$'s are positive, this yields $a(L_1) = a(L_2)$.

No other solutions were received.

### The *n*th Mean Value Theorem

E 3214 [1987, 548]. *Proposed by Guo Qiang Zhang, University of Cambridge, England.*

Let $f$ be a real function with $n + 1$ derivatives on $[a, b]$. Suppose $f^{(i)}(a) = f^{(i)}(b) = 0$ for $i = 0, 1, \ldots, n$. Prove that there is a number $\xi$ in $(a, b)$ such that $f^{(n+1)}(\xi) = f(\xi)$.

$G'$ yields a number $c$ between $x_1$ and $x_2$ such that $G'(c) = 0$, i.e., $f^{(n+1)}(c) = f(c)$. (For the Intermediate Value Theorem for Derivatives see R. P. Boas, Jr., *A Primer of Real Functions*, §21.) □

*Editorial comment.* 25 solvers obtained the result by applying Rolle's Theorem to $h(x) = e^{-x}\sum_{i=0}^{n} f^{(i)}(x)$. B. G. Klein and A. L. Holshouser jointly obtained the same stronger result as Levine. Several generalizations were supplied. M. Falkowitz proved that whenever a polynomial $c_0 + c_1 r + \cdots c_{n+1} r^{n+1}$ has a real root, the given conditions on $f$ imply a solution to $c_0 f(z) + \cdots + c_{n+1} f^{(n+1)}(z) = 0$ in $(a, b)$. T. Jager and D. E. Manes (independently) obtained a special case of this generalization. M. S. Klamkin and A. Meir proved that, if $f$ satisfies the conditions of the problem and $g$ has $n+1$ derivatives on $[a, b]$, then there is a solution to $g(z) f^{(n+1)}(z) = (-1)^{n+1} g^{(n+1)}(z) f(z)$ in $(a, b)$.

The continuity and differentiability assumption given in the first sentence of the problem can obviously be weakened to the assumption that $f, f', \ldots, f^{(n-1)}, f^{(n)}$ are continuous on the closed interval $[a, b]$ and that $f^{(n+1)}$ exists on the open interval $(a, b)$. However, Levine remarks that we can get by with the following still weaker assumption:

(i) $f$ is continuous on the closed interval $[a, b]$,

(ii) $f', f'', \ldots, f^{(n)}$ are continuous on the half-open interval $[a, b)$,

(iii) $f^{(n+1)}$ exists on the open interval $(a, b)$.

To see this we need only note that the proof of Taylor's Theorem given by J. Wolfe, this *Monthly* 60 (1953) 415 gives $(**)$ under the conditions just stated.

Solved also by 34 other readers and the proposer. One incorrect and one partial solution were received.

### Never an Integer

E 3249 [1988, 131]. *Proposed by Toma Albu, University of Bucharest, Romania.*

Show that, for all primes $p$ and all natural numbers $m$, the sum

$$\sum_{k=1}^{m} \frac{k}{1 + k(p-1)}$$

is not an integer.

*Solution by Saeid Zakeri (student), University of Tehran, Iran.* Let $A_n = (p^n - 1)/(p - 1)$. We note first that when $k = A_n$ the $k$th denominator $1 + k(p-1)$ equals $p^n$, while for $A_n < k < A_{n+1}$ the $k$th denominator is not divisible by $p^n$. To see the second statement, suppose that for some $A_n < k < A_{n+1}$ we have $1 + k(p-1) = hp^n$, where necessarily $2 \leqslant h \leqslant p-1$. This can be rewritten as $k = [(h-1)p^n + p^n - 1]/(p-1)$, which is impossible because $p - 1$ cannot divide $(h-1)p^n$.

This being so, the maximal power $p^n$ divides exactly one denominator, that of the $k$th term where $k = A_n$ and $A_n \leqslant m < A_{n+1}$. Since $A_n \equiv 1 \pmod{p}$, the terms of the sum can be combined into a single rational fraction whose numerator is not divisible by $p$ but whose denominator is divisible by $p^n$.

*Editorial comment.* Stephen Gagola noted that the same argument applies to any sum $\sum_{k=1}^{n} f(k)/(1 + k(p-1))$ where $f : \mathbb{Z}^+ \to \mathbb{Z}^+$ is any function with $f(k) \not\equiv 0$

(mod $p$) whenever $k \equiv 1 \pmod p$. Several readers noted that

$$(p-1)\sum_{k=1}^{m} k/(1+k(p-1)) = m - \sum_{k=1}^{m} 1/(1+k(p-1))$$

and proceeded to show that the right-hand sum cannot be an integer. In general, the sum $\sum_{k=1}^{m} 1/(a+kd)$ cannot be an integer for any positive integers $a$ and $d$. Chris Caldwell gave the references T. Nagell, "Eine Eigenschaft gewissen Summen," *Skrifter Oslo*, 13 (1923) 10–15, and T. Nagell, *Introduction to Number Theory*, Wiley, 1951, p. 125 (Problem 51), and Kee-Wai Lau mentioned a paper in Chinese by Z. Shan in *Shuxue Tongxun* (*Communications in Mathematics*), 1 (1985) 30–31 and 26.

Solved also by S.-J. Bang (Korea), C. K. Caldwell, D. Callan, S. M. Gagola, Jr., F. Gillespie, W. Janous (Austria), H.-S. Ki (Korea), Y. H. H. Kwong, K.-W. Lau (Hong Kong), O. P. Lossers, D. E. Manes, L. E. Mattics, K. McInturff, D. Neuenschwander (Switzerland), A. Pedersen (Denmark), R. Persky, A. J. Schwenk, L. Thurston, the George Washington University Problem Group, and the proposer.

### A Subset Not Generating a Group of Orthogonal Matrices

E 3261 [1988, 351]. *Proposed by Detlef Laugwitz, Technische Hochschule, Darmstadt, West Germany.*

Let $G$ be the group of 3 by 3 orthogonal matrices with rational entries. Let $S$ be the subset of $G$ consisting of the six 3 by 3 permutation matrices and all matrices of the form

$$\begin{pmatrix} a/c & b/c & 0 \\ -b/c & a/c & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

where $a, b, c$ are integers with $a^2+b^2=c^2>0$. Does $S$ generate $G$?

*Solution by Jim Delany, California Polytechnic Institute, San Luis Obispo, CA.* The matrix below is not in the group generated by $S$.

$$\begin{pmatrix} 2/3 & 2/3 & 1/3 \\ 2/3 & -1/3 & -2/3 \\ 1/3 & -2/3 & 2/3 \end{pmatrix}$$

Every matrix in $G$ has the form $M/c$, where $c$ is an integer and $M$ is a matrix with integer entries. For the permutation matrices, $c=1$. For the other matrices in $S$ (with $a^2+b^2=c^2$), we may assume that the fractions $a/c$ and $b/c$ are reduced and that $a, b, c$ is a primitive Pythagorean triple. In that case, $a$ and $b$ are relatively prime. It is known that an integer of the form $a^2+b^2$ with $a$ and $b$ coprime cannot have a prime divisor of the form $4n+3$ [see *An Introduction to the Theory of Numbers*, G. H. Hardy and E. M. Wright, Fourth edition, Oxford, Theorem 367]. Thus $c^2$ can have no such divisor, so neither can $c$.

The denominators of the matrices in $S$ have no prime factors of the form $4n+3$, so the denominators of products of matrices in $S$ cannot have such factors. Therefore the matrix displayed above is not in the group generated by $S$.

*Editorial comment.* Christoph Soland proved that in fact the group generated by $S$ has infinite index in $G$. In a letter to the editor of the MONTHLY [1987, 757] John Cremona described how to obtain all elements of $G$.

Also solved by H. Aboutaleb and P. Daugherty, D. Callan, C. Soland (Switzerland), J. Ferrer (Spain), and the proposer. One incorrect solution was received.

**An Elusive Set of Natural Numbers**

E 3266 [1988, 456]. *Proposed by John C. Turner, University of Waikato, Hamilton, New Zealand.*

Let $\mathbb{N}$ be the set of natural numbers. If $S \subseteq \mathbb{N}$ and $n \in \mathbb{N}$, let $S + n = \{s + n\colon s \in S\}$. Define a sequence of sets $S_1, S_2, S_3, \ldots$ recursively by putting $S_1 = \{1\}$ and $S_k = (S_{k-1} + k) \cup \{2k - 1\}$ for $k = 2, 3, 4, \ldots$. What is $\mathbb{N} - \bigcup_{k=1}^{\infty} S_k$?

*Solution by David M. Wells, Penn State University, New Kensington, PA.* The integers omitted by all $S_k$ are the positive powers of 2. Since $\binom{k}{2} + k = \binom{k+1}{2}$ and $2k - 1 = \binom{k+1}{2} - \binom{k-1}{2}$, it can be shown inductively that, if $k \geqslant 2$,

$$S_k = \left\{\binom{k+1}{2} - \binom{j}{2}\colon 1 \leqslant j < k\right\} = \{(k+j)(k+1-j)/2\colon 1 \leqslant j < k\}.$$

Thus a positive integer $n$ belongs to some $S_k$ if and only if $2n = (k+j)(k+1-j)$ for some $j$, $k$ with $j < k$. The two factors on the right have opposite parity, so $n$ cannot be a power of 2. When $n$ is not a power of 2, there are integers $p \geqslant 2$ and $q \geqslant 2$ of opposite parity such that $2n = pq$. Then $2n = (k+j)(k+1-j)$ for $k = (p + q - 1)/2$ and $j = (|p - q| + 1)/2$.

*Editorial comment.* It follows by induction that $S_k$ is the set of integers expressible as a sum of consecutive integers $j + (j+1) + \cdots + k$ with $j < k$. Thus, as noted by Jerrold Grossman, this problem is equivalent to determining the natural numbers not expressible as a sum of consecutive integers, which is a problem in §1.3 of Niven and Zuckerman's *Introduction to the Theory of Numbers* and has a similar easy solution described by Melfried Olson in "Sequentially So," *Math. Magazine*, 52 (1979) 297–298.

Also solved by 48 other readers and the proposer. Three incorrect solutions were received.

## ADVANCED PROBLEMS

6609. *Proposed by Doug Bowman and Tad White, University of California at Los Angeles.*

For each positive integer $n$ let $g(n)$ be the number of digits greater than 4 in the decimal expansion of $2^n$. Is it true that

$$\sum_{n=1}^{\infty} \frac{g(n)}{2^n} = \frac{2}{9}?$$

6610. *Proposed by A. W. Goodman, University of South Florida, Tampa.*

Suppose the sequence $x_1, x_2, \ldots$ is defined by $x_1 = 1$ and $x_n = x_{n-1} + kx_{n-1}^{-c}$ for $n > 1$, where $k$ and $c$ are given positive constants. Find an asymptotic formula for $x_n$.

6611. *Proposed by O. Hajek and D. Singer, Case Western Reserve University, Cleveland, OH.*

Suppose $X$ is a compact metric space with metric $d$ and suppose $T$: $X \to X$ is a homeomorphism of $X$. If $d(x, y) \geq d(Tx, Ty)$ for all $x$, $y$ in $X$, prove that $T$ is an isometry (i.e., $d(x, y) = d(Tx, Ty)$ for all $x$, $y$ in $X$).

## SOLUTIONS OF ADVANCED PROBLEMS

### A Nonvanishing Determinant

6543 [1987, 387]. *Proposed by A. S. Cavaretta, Jr. and C. R. Selvaraj, Kent State University, Kent, OH.*

Let $m_1, m_2, \ldots, m_r$ be distinct odd integers. Show that the determinant

$$\begin{vmatrix} 1 & 2^{m_1} & 3^{m_1} & 4^{m_1} & 5^{m_1} & \cdots & (r+1)^{m_1} \\ & & \vdots & & & & \vdots \\ 1 & 2^{m_r} & 3^{m_r} & 4^{m_r} & 5^{m_r} & \cdots & (r+1)^{m_r} \\ 1 & 0 & -1 & 0 & 1 & \cdots & d_{r+1} \end{vmatrix}$$

is nonzero, where the $j$th entry in the last row is

$$d_j = \sin(\pi j/2), \qquad j = 1, 2, \ldots, r+1.$$

*Solution by the proposers and T. N. T. Goodman, University of Dundee.* Let $(s_1, \ldots, s_n)$ be a sequence of real numbers. The number of *sign changes* in the sequence, written $S^-(s_1, \ldots, s_n)$, is the number of blocks $(s_j, s_{j+1}, \ldots, s_k)$ in which $s_j s_k < 0$ and $s_i = 0$ for $j < i < k$. Clearly this is unaltered by the arbitrary insertion or deletion of terms equal to zero. Laguerre's generalization of Descartes' Rule of Signs states that, if $m_1 < m_2 < \cdots < m_n$, then the number of positive zeros of the function $s_1 x^{m_1} + s_2 x^{m_2} + \cdots + s_n x^{m_n}$ is bounded above by $S^-(s_1, \ldots, s_n)$. (See Problem 77, Part V in G. Pólya and G. Szegö, *Problems and Theorems in Analysis II*, Springer, 1976.) The generalized Vandermonde matrix is $[x_j^{m_i}]$, where $m_1 < \cdots < m_n$ and $0 < x_1 < \cdots < x_n$. If the rows of this matrix were dependent, there would exist $\{c_i\}$, not all zero, so that $\Sigma c_i x_j^{m_i} = 0$ for $1 \leqslant j \leqslant n$. This contradicts Laguerre's Theorem, because $S^-(c_1, \ldots, c_n) \leqslant n-1$; thus, the generalized Vandermonde has nonzero determinant. (In fact, since it is a continuous function of the $x_i$'s, by letting $x_n \to \infty$ and induction, we see that it is positive.)

We need one other nontrivial fact. The matrix $[a_{ij}]$ is called *totally positive* if every minor has nonnegative determinant. I. J. Schoenberg proved in 1930 that

totally positive matrices have a variation diminishing property. That is, if $s'_j = \Sigma a_{ij} s_i$, then $S^-(s'_1, \dots, s'_n) \leqslant S^-(s_1, \dots, s_n)$. (See S. Karlin, *Total Positivity*, Stanford Univ. Press, 1968, or S. Karlin, "Total positivity and variation diminishing transformations" pp. 269–273 in I. J. Schoenberg, *Selected Papers* Vol. 2, Birkhäuser, Boston, 1988.) We shall need one of the first totally positive matrices introduced by Schoenberg: $a_{ij} = \binom{m_i}{j}$, where $m_1 < m_2 < \cdots < m_n$.

Let $A$ be the matrix as given in the problem and suppose that $\det A = 0$, so the rows are dependent. The argument of the first paragraph shows that the first $r$ rows of $A$ are independent, and so the last row is a linear combination of the first $r$. In other words, there exist $\{c_{m_i}\}$ so that $\Sigma_1^r c_{m_i} j^{m_i} = d_j$, $1 \leqslant j \leqslant r+1$. (Define $c_k = 0$ if $k$ is odd and $k$ is not one of the $m_i$'s.) Define the polynomials $\varphi(x) = \Sigma_1^r c_{m_i} x^{m_i}$ and $\psi(x) = \varphi(x+1) + \varphi(x-1)$, so that $\varphi(j) = d_j$ for $1 \leqslant j \leqslant r+1$ and $\psi(j) = d_{j-1} + d_{j+1} = 0$ for $1 \leqslant j \leqslant r$. (It is easy to see that $\varphi(0) = 0 = d_0$.) We have

$$\psi(x) = \varphi(x+1) + \varphi(x-1) = \sum_{i=1}^{r} c_{m_i}\left((x+1)^{m_i} + (x-1)^{m_i}\right)$$

$$= 2\sum_{i=1}^{r} c_{m_i}\left\{\sum_{k \text{ odd}} \binom{m_i}{k} x^k\right\} = 2\sum_{k \text{ odd}}\left\{\sum_{i=1}^{r} c_{m_i}\binom{m_i}{k}\right\} x^k := \sum_{k \text{ odd}} c'_k x^k.$$

It is easy to see that $\psi(x)$ has degree equal to that of $\varphi(x)$. By Schoenberg's theorem on variation diminishing transformations and the suitable deletion of zero terms, we see that $S^-(c'_1, \dots, c'_{m_r}) \leqslant S^-(c_1, \dots, c_{m_r}) = S^-(c_{m_1}, \dots, c_{m_r}) \leqslant r-1$; but $\psi(x)$ has $r$ positive zeros, and this violates Laguerre's Theorem.

Also solved by L. E. Mattics.

### A Function Induced by the Cantor Set

6554 [1987, 799]. *Proposed by Boo Rim Choe (student), University of Wisconsin, Madison.*

Let $E$ be the Cantor set. It is well-known that $E + E = [0,2]$, where $E + E = \{x + y \colon x \in E, y \in E\}$. (Cf. R. P. Boas, Jr., *A Primer of Real Functions*, §20.) Define $h \colon [0,1] \to [0,1]$ as follows:

$$h(s) = \sup\{y \colon y + x = 2s;\ x, y \in E\}.$$

(1) Prove that $h$ is a Borel function and compute its mean value.
(2) Given $s \in E$, determine the value of $h(s)$ in terms of its ternary expansion.
(3) Find all points of continuity of $h$.
(4) Prove that $h$ is not of bounded variation.

*Solution by O. P. Lossers, University of Technology, Eindhoven, The Netherlands.* We shall show the following. Let $s$ have the ternary expansion $0.s_1 s_2 s_3 \dots$, in which the "tails" $0222\dots$ and $2000\dots$ are forbidden, i.e., are replaced by $1000$ and $1222$, respectively. Then $h(s)$ has the ternary expansion $0.z_1 z_2 z_3$, where $z_i = 0$ if $s_i = 0$ and $z_i = 2$ if $s_i \neq 0$. Thus $h(s) \in E$. Note that, in the ternary expansions obtained for $h(s)$, the tails $10000\dots$ and $12222\dots$ do *not* occur.

First suppose that $1/3 < s < 2/3$. Observe that

$$x \in E \quad \text{and} \quad x \leqslant \frac{1}{3}$$

totally positive matrices have a variation diminishing property. That is, if $s_j' = \Sigma a_{ij}s_i$, then $S^-(s_1', \dots, s_n') \leqslant S^-(s_1, \dots, s_n)$. (See S. Karlin, *Total Positivity*, Stanford Univ. Press, 1968, or S. Karlin, "Total positivity and variation diminishing transformations" pp. 269–273 in I. J. Schoenberg, *Selected Papers* Vol. 2, Birkhäuser, Boston, 1988.) We shall need one of the first totally positive matrices introduced by Schoenberg: $a_{ij} = \binom{m_i}{j}$, where $m_1 < m_2 < \cdots < m_n$.

Let $A$ be the matrix as given in the problem and suppose that $\det A = 0$, so the rows are dependent. The argument of the first paragraph shows that the first $r$ rows of $A$ are independent, and so the last row is a linear combination of the first $r$. In other words, there exist $\{c_{m_i}\}$ so that $\Sigma_1^r c_{m_i} j^{m_i} = d_j$, $1 \leqslant j \leqslant r+1$. (Define $c_k = 0$ if $k$ is odd and $k$ is not one of the $m_i$'s.) Define the polynomials $\varphi(x) = \Sigma_1^r c_{m_i} x^{m_i}$ and $\psi(x) = \varphi(x+1) + \varphi(x-1)$, so that $\varphi(j) = d_j$ for $1 \leqslant j \leqslant r+1$ and $\psi(j) = d_{j-1} + d_{j+1} = 0$ for $1 \leqslant j \leqslant r$. (It is easy to see that $\varphi(0) = 0 = d_0$.) We have

$$\psi(x) = \varphi(x+1) + \varphi(x-1) = \sum_{i=1}^{r} c_{m_i}\left((x+1)^{m_i} + (x-1)^{m_i}\right)$$

$$= 2\sum_{i=1}^{r} c_{m_i}\left\{\sum_{k\text{ odd}} \binom{m_i}{k} x^k\right\} = 2\sum_{k\text{ odd}}\left\{\sum_{i=1}^{r} c_{m_i}\binom{m_i}{k}\right\} x^k := \sum_{k\text{ odd}} c_k' x^k.$$

It is easy to see that $\psi(x)$ has degree equal to that of $\varphi(x)$. By Schoenberg's theorem on variation diminishing transformations and the suitable deletion of zero terms, we see that $S^-(c_1', \dots, c_{m_r}') \leqslant S^-(c_1, \dots, c_{m_r}) = S^-(c_{m_1}, \dots, c_{m_r}) \leqslant r-1$; but $\psi(x)$ has $r$ positive zeros, and this violates Laguerre's Theorem.

Also solved by L. E. Mattics.

### A Function Induced by the Cantor Set

6554 [1987, 799]. *Proposed by Boo Rim Choe (student), University of Wisconsin, Madison.*

Let $E$ be the Cantor set. It is well-known that $E + E = [0,2]$, where $E + E = \{x + y\colon x \in E, y \in E\}$. (Cf. R. P. Boas, Jr., *A Primer of Real Functions*, §20.) Define $h\colon [0,1] \to [0,1]$ as follows:

$$h(s) = \sup\{y\colon y + x = 2s;\ x, y \in E\}.$$

(1) Prove that $h$ is a Borel function and compute its mean value.
(2) Given $s \in E$, determine the value of $h(s)$ in terms of its ternary expansion.
(3) Find all points of continuity of $h$.
(4) Prove that $h$ is not of bounded variation.

*Solution by O. P. Lossers, University of Technology, Eindhoven, The Netherlands.* We shall show the following. Let $s$ have the ternary expansion $0.s_1s_2s_3\dots$, in which the "tails" $0222\dots$ and $2000\dots$ are forbidden, i.e., are replaced by $1000$ and $1222$, respectively. Then $h(s)$ has the ternary expansion $0.z_1z_2z_3$, where $z_i = 0$ if $s_i = 0$ and $z_i = 2$ if $s_i \neq 0$. Thus $h(s) \in E$. Note that, in the ternary expansions obtained for $h(s)$, the tails $10000\dots$ and $12222\dots$ do *not* occur.

First suppose that $1/3 < s < 2/3$. Observe that

$$x \in E \quad\text{and}\quad x \leqslant \frac{1}{3}$$

We remark further that $h$ is not differentiable at any point of $(0,1)$. To see this, consider first an $s$ with a ternary expansion that contains infinitely many zeros but does not consist solely of zeros from some point on. Then, if the $N$th digit is a zero,

$$\frac{h(s+3^{-N})-h(s)}{3^{-N}}=2, \qquad \frac{h(s+2\cdot 3^{-N})-h(s)}{2\cdot 3^{-N}}=1.$$

If $s$ has a ternary expansion with infinitely many ones, then, if the $N$th digit is a one,

$$\frac{h(s+3^{-N})-h(s)}{3^{-N}}=0, \qquad \frac{h(s-3^{-N})-h(s)}{-3^{-N}}=2.$$

If the ternary expansion of $s$ contains infinitely many twos but does not consist solely of twos from some point on, then, if the $N$th digit is a two,

$$\frac{h(s-3^{-N})-h(s)}{3^{-N}}=0, \qquad \frac{h(s-2\cdot 3^{-N})-h(s)}{-2\cdot 3^{-N}}=1.$$

Thus in each of the above three cases $h$ is not differentiable at the point $s$. In the remaining cases the ternary expansion of $s$ has a tail $10000\dots$ or a tail $12222\dots$ and then $s$ is a point where $h$ has already been proved to be discontinuous.

Solved also by Abraham Shimron (Israel) and the proposer.

### A Tangential Series

6560 [1987, 1010]. *Proposed by Ambati Jaya Krishna (student), John Hopkins University, Baltimore, MD, A. Murali Mohan Rao, Iona College, New Rochelle, NY, and Gomathi S. Rao, Baltimore, MD.*

If $x$ and $y$ are odd positive integers, evaluate

$$\sum_{n=1}^{\infty}\frac{1}{n^2}\tan\left(\frac{n\pi}{x}\right)\tan\left(\frac{n\pi}{y}\right).$$

*Solution by Stan Philipp, Penn State University (Altoona Campus).* Let $S(x, y)$ denote the given sum and $g$ the greatest common divisor of $x$ and $y$. We shall show that

$$S(x,y)=\pi^2(g^2-1)/(2xy).$$

Define

$$f_x(t)=(-1)^{[x|t|/\pi]}, \qquad -\pi<t<\pi,$$

where here and henceforth the square brackets shall denote the greatest integer function. Observe that $f_x$ is even, and constant on each interval

$$\frac{j\pi}{x}<t<\frac{(j+1)\pi}{x}, \qquad j=0,1,\dots,x-1.$$

Since for odd integers $x$ we have

$$\sum_{j=0}^{x-1}(-1)^j\cos(2j+1)s=(1+\cos 2xs)/(2\cos s),$$

We remark further that $h$ is not differentiable at any point of $(0,1)$. To see this, consider first an $s$ with a ternary expansion that contains infinitely many zeros but does not consist solely of zeros from some point on. Then, if the $N$th digit is a zero,

$$\frac{h(s+3^{-N})-h(s)}{3^{-N}}=2,\qquad \frac{h(s+2\cdot 3^{-N})-h(s)}{2\cdot 3^{-N}}=1.$$

If $s$ has a ternary expansion with infinitely many ones, then, if the $N$th digit is a one,

$$\frac{h(s+3^{-N})-h(s)}{3^{-N}}=0,\qquad \frac{h(s-3^{-N})-h(s)}{-3^{-N}}=2.$$

If the ternary expansion of $s$ contains infinitely many twos but does not consist solely of twos from some point on, then, if the $N$th digit is a two,

$$\frac{h(s-3^{-N})-h(s)}{3^{-N}}=0,\qquad \frac{h(s-2\cdot 3^{-N})-h(s)}{-2\cdot 3^{-N}}=1.$$

Thus in each of the above three cases $h$ is not differentiable at the point $s$. In the remaining cases the ternary expansion of $s$ has a tail $10000\ldots$ or a tail $12222\ldots$ and then $s$ is a point where $h$ has already been proved to be discontinuous.

Solved also by Abraham Shimron (Israel) and the proposer.

### A Tangential Series

6560 [1987, 1010]. *Proposed by Ambati Jaya Krishna (student), John Hopkins University, Baltimore, MD, A. Murali Mohan Rao, Iona College, New Rochelle, NY, and Gomathi S. Rao, Baltimore, MD.*

If $x$ and $y$ are odd positive integers, evaluate

$$\sum_{n=1}^{\infty}\frac{1}{n^2}\tan\left(\frac{n\pi}{x}\right)\tan\left(\frac{n\pi}{y}\right).$$

*Solution by Stan Philipp, Penn State University (Altoona Campus).* Let $S(x,y)$ denote the given sum and $g$ the greatest common divisor of $x$ and $y$. We shall show that

$$S(x,y)=\pi^2(g^2-1)/(2xy).$$

Define

$$f_x(t)=(-1)^{[x|t|/\pi]},\qquad -\pi<t<\pi,$$

where here and henceforth the square brackets shall denote the greatest integer function. Observe that $f_x$ is even, and constant on each interval

$$\frac{j\pi}{x}<t<\frac{(j+1)\pi}{x},\qquad j=0,1,\ldots,x-1.$$

Since for odd integers $x$ we have

$$\sum_{j=0}^{x-1}(-1)^j\cos(2j+1)s=(1+\cos 2xs)/(2\cos s),$$

it is not hard to obtain the Fourier coefficient formula

$$\hat{f}_x(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f_x(t)e^{-int}\,dt = \begin{cases} 0 & n \text{ odd} \\ \dfrac{1}{x} & n = 0 \\ \dfrac{2}{n\pi}\tan\dfrac{n\pi}{2x} & n \neq 0 \text{ and } n \text{ even.} \end{cases}$$

By Parseval's theorem,

$$\frac{1}{\pi}\int_0^{\pi} f_x(t)f_y(t)\,dt = \sum_{n=-\infty}^{\infty} \hat{f}_x(n)\hat{f}_y(n).$$

Using the change of variable $t = \pi s/y$ on the left and the Fourier coefficient formula on the right, we obtain

$$F(x, y) := \frac{1}{y}\int_0^y (-1)^{[sx/y]+[s]}\,ds = \frac{1}{xy} + \frac{2}{\pi^2}S(x, y).$$

It now suffices to show

$$F(x, y) = g^2/(xy).$$

Assume (without loss of generality) $x \leqslant y$ and set

$$a_i = \left[\frac{iy}{x}\right], \qquad i = 0, 1, \dots, x.$$

Note that $a_x = y$ is odd, and

$$a_i \leqslant \frac{iy}{x} < 1 + a_i \leqslant a_{i+1}, \qquad i = 0, 1, \dots, x-1.$$

Let $\phi(s)$ denote the integrand of $yF(x, y)$. Then

$$yF(x, y) = \sum_{i=0}^{x-1}\int_{a_i}^{a_{i+1}}\phi(s)\,ds = \sum_{i=0}^{x-1}\int_{a_i}^{1+a_i}\phi(s)\,ds + \sum_{i=0}^{x-1}\int_{1+a_i}^{a_{i+1}}\phi(s)\,ds$$
$$= \sum_1 + \sum_2.$$

The second sum $\Sigma_2$ vanishes, since it equals

$$\sum_{i=0}^{x-1}(-1)^i\int_{1+a_i}^{a_{i+1}}(-1)^{[s]}\,ds = \sum_{i=0}^{x-1}(-1)^{i+1+a_i}\{1 + (-1)^{a_{i+1}-a_i}\}/2$$
$$= \sum_{i=0}^{x-1}\frac{(-1)^{i+1+a_{i+1}} - (-1)^{i+a_i}}{2}$$
$$= \frac{(-1)^{x+a_x} - (-1)^{0+a_0}}{2} = 0.$$

Next,

$$\sum_1 = \sum_{i=0}^{x-1}\left(\int_{a_i}^{iy/x}\phi(s)\,ds + \int_{iy/x}^{1+a_i}\phi(s)\,ds\right)$$

$$= \sum_{i=0}^{x-1}\left\{(-1)^{i-1+a_i}\left(\frac{iy}{x} - a_i\right) + (-1)^{i+a_i}\left(1 + a_i - \frac{iy}{x}\right)\right\}$$

$$= \frac{1}{x}\sum_{i=0}^{x-1}(-1)^{iy-a_ix+1}(2iy - 2a_ix - x).$$

Now the collection of $x$ numbers

$$iy - a_ix, \qquad i = 0,1,\dots,x-1,$$

consists of the numbers

$$kg, \qquad k = 0,1,\dots,\frac{x}{g} - 1,$$

with each number $kg$ occurring $g$ times; the reason for this is that $iy - a_ix$ is the least nonnegative residue of $iy$ modulo $x$. Hence

$$\sum_1 = \frac{1}{x}\sum_{k=0}^{x/g-1} g(-1)^{kg+1}(2kg - x)$$

$$= \frac{2g^2}{x}\sum_{k=0}^{x/g-1}(-1)^{k+1}k - g\sum_{k=0}^{x/g-1}(-1)^{k+1}$$

$$= \frac{2g^2}{x}\,\frac{\left(1 - \dfrac{x}{g}\right)}{2} - g(-1) = g^2/x,$$

and so $yF(x, y) = \Sigma_1 = g^2/x$. Thus the assertion of the problem is established.

*Editorial comment.* The proposers gave a somewhat more symmetrical solution based on

$$\int_0^1(-1)^{[tx]}(-1)^{[ty]}\,dt = \frac{g^2}{xy}$$

for $x$ and $y$ odd. L. E. Mattics employed the Fourier expansion of $t(\pi - t)$ and properties of roots of unity; he also established the identity

$$g^2 + \sum_{i=0}^{xy-1}(-1)^{[i/y]+[(-1-i)/x]} = 0.$$

A. Pedersen (Denmark) and I. E. Leonard (Canada) gave proofs based on contour integration. Leonard established

$$S(x, y) = \frac{\pi^2}{2p^2}\sum_{n=1}^{p-1}\frac{\tan\dfrac{n\pi}{x}\tan\dfrac{n\pi}{y}}{\sin^2\dfrac{n\pi}{p}}$$

where $p$ is the least common multiple of $x$ and $y$, and cited the reference B. C. Berndt, The evaluation of infinite series by contour integration, *Univ. Beograd Publ. Elektrotehn. Fak. Ser. Mat. Fiz.*, No. 435 (1973) 119–122.

EDITED BY JOSEPH KONHAUSER
*Department of Mathematics, Macalester College, St. Paul, MN 55105*


*The Pólya Picture Album—Encounters of a Mathematician.* Edited by Gerald L. Alexanderson. Birkhäuser Boston, 1987, 140 pp. ISBN 0-8176-3352-9, $35.00.


FRANK HARARY
*Department of Computer Science, New Mexico State University, Las Cruces, NM 88003*


George Pólya was my hero both as a mathematician and as a teacher. His greatest combinatorial work was the classic 110-page paper [**8**]. Its "Hauptsatz" is the "Pólya Enumeration Theorem" (PET). The PET enabled me (often with gifted coauthors) to count many types of finite graph structures by specifying the permutation group which characterized equivalent configurations. Thus it was that the two binary operations on permutation groups $A$ and $B$ called the power group $B^A$ and the exponentiation group $[B]^A$ were discovered. The "Power Group Enumeration Theorem" of [**7**] is entirely equivalent to de Bruijn's earlier generalization [**2**] of the PET since it shows that the PET applied to the power group gives de Bruijn's formula. The exponentiation group [**4**] characterizes the automorphisms of the complete bipartite graph $K_{n,n}$ and the hypercube graph $Q_n$, amongst others.

I first met Professor Pólya by mail while I was trying to determine the number of different graphs with a given number of nodes and edges. Luckily I did not learn until later that each of the following mathematicians had already independently solved this problem, not knowing that Pólya himself had already done it: R. Davis, E. Gilbert, A. Gleason, D. Slepian. At that time nobody knew that Redfield [**10**] had already published this result, using extremely obscure notation and terminology. I shall never forget Pólya's reply to my letter asking him for the formula for the number of graphs which I heard (from Hans Samelson) he had found. Approximately, he wrote:

> Dear Harary,
>
> Let $g_{p,q}$ be the number of graphs with $p$ points and $q$ lines and define the counting polynomial
>
> $$g_p(x) = \sum_{q=0}^{p(p-1)/2} g_{p,q}x^q.$$
>
> Then $g_p(x) = \cdots$ (a formula which covered the next page).

He then carefully and clearly illustrated the result for $p = 4$.

After I finally understood both the group involved (the pair group of the symmetric group) and the PET (while working with the great theoretical physicist George Uhlenbeck), I was able to count directed graphs, rooted graphs, and connected graphs. I then wrote to Professor Pólya and asked him respectfully if I could submit these results as a joint paper with him. He wrote as follows, and of

course his wishes were followed in [3]:

> Dear Harary,
>
> I cannot possibly be a coauthor of the paper for I would then have to read it and that takes time from writing books. But please do me two favors:
>
> 1. From now on say "digraph" instead of "directed graph." It is clear and short and it will catch on (it did).
>
> 2. I agree that I deserve more than a footnote so please mention me in the text.

This correspondence led to his inviting me to spend two weeks in Stanford in 1960, giving ten talks surveying graph theory, all of which he attended. He asked me then if I could find a formula for the determinant of the adjacency matrix of a graph. I could, having derived it a few years earlier, at which time I felt that it was not deep enough for publication. He told me that it may be easy to prove when you look at it *just right* but it is new and interesting so please publish it, which I did [5]. He explained his interest by showing me the problem in his *opus magnum* with Szegö [9] in which he asked this question for the five platonic graphs.

During that visit and several later ones, he showed me, at his home, with the greatest of enthusiasm and pleasure, many of the photographs which appear in this wonderful book, so carefully and beautifully edited by one of his students, Gerry Alexanderson.

Many of the greatest mathematicians of the century appear in this picture album. The reader will have the unique opportunity of seeing the faces of many very famous mathematicians, including: Bernays, Besicovich, Bierberbach, Bohr, Carathéodory, Cartan, Cramér, Davenport, Dieudonné, Einstein, Erdélyi, Errera, Erdös, Fekete, Fejér, Hadamard, Hardy, Hausdorff, Hilbert, Hille, Hopf, Klein, Kuratowski, Landau, Lehmer (Derrick and Emma), Littlewood, Menger, Minkowski, Mittag-Leffler, Nevanlinna, Neyman, Noether, Pauli, Perron, Plancherel, Poincaré, Ostrowski, Rademacher, Reidemeister, Rényi, Robinson (Julia and Raphael), Runge, Schur, Sierpinski, Spencer, Szász, Szegö, Tamarkin, Uspensky, Weyl, Wiener, and Zygmund.

The humanity, happiness, friendliness, sense of humor, and genuine humility of George Pólya can be seen in the following selected quotations:

> We start off with a picture of Einstein. Einstein was then young and good looking, not the Einstein we usually see.
>
> My connection with Hurwitz was deeper and my debt to him greater than to any other colleague. We had a special way we worked. I would visit him and we would sit in his study and talk mathematics until he finished his cigar.
>
> Hilbert once had a student who stopped coming to his lectures, and was finally told that the young man had gone off to become a poet. Hilbert is reported to have remarked: "I never thought he had enough imagination to be a mathematician." Hilbert and his wife were giving a party at their house when Mrs. Hilbert noted that he had failed to put on a fresh shirt, so

> enthusiasm, the twinkle in his eye, his tremendous curiosity, his generosity with his time, his spry energetic walk, his warm genuine friendliness, his welcoming visitors into his home and showing them his pictures of great mathematicians he has known—these are all components of his happy personality. As a mathematician, his depth, speed, brilliance, versatility, power, and universality are all inspiring. Would that there were a way of teaching and learning these traits!

The short biography of Pólya which begins this picture album was written not only masterfully but with love and respect permeating the narrative. Alexanderson is to be warmly congratulated. I recommend this book strongly to all mathematicians, research scholars, professors, teachers, students, and amateurs including all readers of any periodical whose name includes a word such as: science, scientific, scientist, etc.

I encountered just one misprint in the entire book. Even when I told a colleague the page number, he could not find it!

REFERENCES

1. G. L. Alexanderson and L. H. Lange, Obituary of George Pólya, *Bull. London Math. Soc.*, 19 (1987) 559–608.
2. N. G. de Bruijn, Generalization of Pólya's fundamental theorem in enumeration combinatorial analysis, *Indagationes Math.*, 21 (1959) 59–69.
3. F. Harary, The number of linear, directed, rooted, and connected graphs, *Trans. Amer. Math. Soc.*, 78 (1955) 445–463.
4. ______, Exponentiation of permutation groups, *Amer. Math. Monthly*, 66 (1959) 572–575.
5. ______, The determinant of the adjacency matrix of a graph, *SIAM Review*, 4 (1962) 202–210.
6. ______, Homage to George Pólya, *J. Graph Theory*, 1 (1977) 289–290.
7. F. Harary and E. M. Palmer, The power group enumeration theorem, *J. Combinatorial Theory*, 1 (1966) 157–173.
8. G. Pólya, Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen, *Acta Math.*, 68 (1937) 145–254.
9. G. Pólya and G. Szegö, Aufgaben und Lehrsätze aus der Analysis, 2 vol. Springer, Berlin (1925).
10. J. H. Redfield, The theory of group-reduced distributions, *Amer. J. Math.*, 49 (1927) 433–455.

*Geometric Measure Theory. A Beginner's Guide.* By Frank Morgan. Academic Press, 1988. vii + 145 pp.

FREDERICK J. ALMGREN, JR
*Department of Mathematics, Princeton University, Princeton, NJ 08544*

*"Find the surface of least area spanning a given contour."*

This seemingly innocent problem can bring visions of iridescent soap films or perhaps dazzling computer renderings of minimal surfaces. Making mathematics of it, however, opens a Pandora's box of questions, e.g., What is a surface?, What is the area of a surface?, What does it mean for a surface to span a contour?, What does it mean to "find" a surface? A soap film, for example, will form on a piece of wire tied into a simple overhand knot whether or not the ends of the wire are joined together. In this case the notions of surface and of spanning need to be adapted from the underlying physical problem rather than more common mathematical devices.

> enthusiasm, the twinkle in his eye, his tremendous curiosity, his generosity with his time, his spry energetic walk, his warm genuine friendliness, his welcoming visitors into his home and showing them his pictures of great mathematicians he has known—these are all components of his happy personality. As a mathematician, his depth, speed, brilliance, versatility, power, and universality are all inspiring. Would that there were a way of teaching and learning these traits!

The short biography of Pólya which begins this picture album was written not only masterfully but with love and respect permeating the narrative. Alexanderson is to be warmly congratulated. I recommend this book strongly to all mathematicians, research scholars, professors, teachers, students, and amateurs including all readers of any periodical whose name includes a word such as: science, scientific, scientist, etc.

I encountered just one misprint in the entire book. Even when I told a colleague the page number, he could not find it!

REFERENCES

1. G. L. Alexanderson and L. H. Lange, Obituary of George Pólya, *Bull. London Math. Soc.*, 19 (1987) 559–608.
2. N. G. de Bruijn, Generalization of Pólya's fundamental theorem in enumeration combinatorial analysis, *Indagationes Math.*, 21 (1959) 59–69.
3. F. Harary, The number of linear, directed, rooted, and connected graphs, *Trans. Amer. Math. Soc.*, 78 (1955) 445–463.
4. ______, Exponentiation of permutation groups, *Amer. Math. Monthly*, 66 (1959) 572–575.
5. ______, The determinant of the adjacency matrix of a graph, *SIAM Review*, 4 (1962) 202–210.
6. ______, Homage to George Pólya, *J. Graph Theory*, 1 (1977) 289–290.
7. F. Harary and E. M. Palmer, The power group enumeration theorem, *J. Combinatorial Theory*, 1 (1966) 157–173.
8. G. Pólya, Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen, *Acta Math.*, 68 (1937) 145–254.
9. G. Pólya and G. Szegö, Aufgaben und Lehrsätze aus der Analysis, 2 vol. Springer, Berlin (1925).
10. J. H. Redfield, The theory of group-reduced distributions, *Amer. J. Math.*, 49 (1927) 433–455.

*Geometric Measure Theory. A Beginner's Guide.* By Frank Morgan. Academic Press, 1988. vii + 145 pp.

FREDERICK J. ALMGREN, JR
*Department of Mathematics, Princeton University, Princeton, NJ 08544*

*"Find the surface of least area spanning a given contour."*

This seemingly innocent problem can bring visions of iridescent soap films or perhaps dazzling computer renderings of minimal surfaces. Making mathematics of it, however, opens a Pandora's box of questions, e.g., What is a surface?, What is the area of a surface?, What does it mean for a surface to span a contour?, What does it mean to "find" a surface? A soap film, for example, will form on a piece of wire tied into a simple overhand knot whether or not the ends of the wire are joined together. In this case the notions of surface and of spanning need to be adapted from the underlying physical problem rather than more common mathematical devices.

Since these questions can have a variety of different answers, the problem of least area turns out to be a collection of problems with many similarities but also with major differences. Some feel at this point that "I don't really want to know any more if it is that complicated." For others, Frank Morgan's guide can help open new mathematical vistas and point the way to beautiful theorems in the geometric calculus of variations with a variety of physical and mathematical applications.

Geometric measure theory acquired its name from the title of Federer's treatise [F] (which would have been called Geometric Integration Theory had not Whitney already published a book of that title). The field is much larger than is suggested by least area problems. Federer's celebrated co-area formula (a curvilinear form of Fubini's theorem), for example, has little to do directly with minimal surfaces but is absolutely fundamental knowledge for the working geometric analyst. It is the experience of the reviewer that many who recognize the power of the field's methods and would like to use them in their own work have been daunted at trying to read Federer's book (it is almost 700 pages long with no illustrations). Morgan has borrowed the title of the treatise, but subtitles his book, "A Beginner's Guide." This is an accurate description. Morgan's book is devoted largely to giving the facts (generally without proofs) and motivation along with lavish geometric illustrations. Geometric measure theory has a strong intuitive and geometric component, much of which has been captured by Morgan. Morgan, incidentally, has become virtually a legend in his own time as a teacher and mathematical expositor.

There is much left to be done in understanding even least area problems. For example, one of the abstract existence and regularity results discussed by Morgan implies that any smooth simple closed curve $\mathscr{C}$ in $\boldsymbol{R}^3$ must be the boundary of a smooth minimal submanifold $\mathscr{M}$ of $\boldsymbol{R}^3$ whose area does not exceed the area of any other manifold (embedded or immersed, oriented or unoriented) having $\mathscr{C}$ as boundary. However, except in special cases, there is no way known to compute such an $\mathscr{M}$ or even reliably estimate what its area should be (possible nonorientability is the most serious difficulty).

Since Morgan's book is called *Geometric Measure Theory* one might ask what measures have to do with surfaces of least area. In retrospect, the field of geometric measure theory during the first half of this century was largely devoted to the development of a theory of measure and integration for $k$-dimensional surfaces in $n$-dimensional spaces when these surfaces might have essential singularities, or might even be Cantor type sets themselves. Incidentally, one now commonly accepts Hausdorff's $k$-dimensional measure in $\boldsymbol{R}^n$ as the definition of surface measure.

There is then a strange twist. One answers the question, "What is a surface?" by saying that a $k$-dimensional surface in $\boldsymbol{R}^n$ is a special kind of measure over all of $\boldsymbol{R}^n$. The measure $\|\mathscr{M}\|$ which corresponds to a $k$-dimensional submanifold $\mathscr{M}$ of $\boldsymbol{R}^n$ assigns to a subset $A$ of $\boldsymbol{R}^n$ Hausdorff's $k$-dimensional measure of the intersection of $\mathscr{M}$ with $A$. Thus the restriction of $\|\mathscr{M}\|$ to $\mathscr{M}$ is the usual surface area measure of $\mathscr{M}$. Surfaces with complicated singularities become measures in the same way. In fact, anything one could ever want to be surface of finite area turns out to be representable by such a measure.

This identification of surfaces with measures over all of $\boldsymbol{R}^n$ provides a way to obtain existence theorems for area minimizing surfaces $S$ spanning a contour $C$. Suppose that $S_1, S_2, S_3, \ldots$ are surfaces spanning $C$ whose areas approach the infimum of areas of all possible surfaces, and let $\|S_1\|, \|S_2\|, \|S_3\|, \ldots$ be the associ-

ated measures. Spaces of measures have very strong compactness properties in the weak topology and, in the present case, there will always exist a measure $\mu$ over $\boldsymbol{R}^3$ which is the weak limit of a subsequence of the $\|S_j\|$'s. This measure is supposed to be a surface of least area spanning $C$. The real work comes, of course, in showing that $\mu = \|S\|$ for a nice surface $S$.

In order to characterize surfaces measure theoretically one is led to introduce measure theoretic analogues of smooth submanifolds and smooth mappings. According to Morgan: "*Geometric Measure Theory* could be described as differential geometry generalized through measure theory to deal with maps and surfaces that are not necessarily smooth, and applied to the calculus of variations."

It is one of the great triumphs of the subject that everything one could reasonably hope for turns out to be true if submanifolds are replaced by "rectifiable sets," smooth maps are replaced by "Lipschitz maps," and everywhere is replaced by "almost everywhere." For example, unions and intersections of rectifiable sets remain rectifiable, and the Lipschitz image of a rectifiable set is rectifiable.

A key first step in recovering $S$ from $\mu$ lies in the structure theorem for sets of finite Hausdorff measure. In the present context this theorem implies that any part of the set carrying $\mu$ which is not rectifiable must be so full of holes that it could contribute nothing to an area minimization problem for spanning surfaces—hence it wasn't there.

This is only part of the story. Even knowing that $\mu = \|S\|$ for a rectifiable $S$ still is not really satisfactory since rectifiable sets sometimes are pretty wild. A substantial part of the research in geometric measure theory in the modern period (since 1960) has been devoted to showing that area minimizing $S$'s are smooth minimal submanifolds except for small singular sets.

One of the central devices of modern geometric measure theory has been to regard oriented surfaces as currents, i.e., as continuous linear functionals on vector spaces of smooth differential forms—this is quite analogous to identifying functions with distributions. Part of how this came about is the following. A fixed differential $n - k$ form $\sigma$ on a compact $n$-dimensional Riemannian manifold $\mathscr{N}$ induces a linear functional $[\![\sigma]\!]$ on the vector space of differential $k$ forms on $\mathscr{N}$ which maps a $k$ form $\omega$ to the number $[\![\sigma]\!](\omega) = \int_{\mathscr{N}}\omega \wedge \sigma$; on the cohomology level this is the pairing which induces Poincaré duality. A smooth oriented $k$-dimensional submanifold $\mathscr{S}$ of $\mathscr{N}$ also induces a linear mapping $[\![\mathscr{S}]\!]$ which maps a form $\omega$ to the number $[\![\mathscr{S}]\!](\omega) = \int_{\mathscr{S}}\omega$. Suppose a $\mathscr{S}$ without boundary is given. Can one find a closed $\sigma$ such that $[\![\sigma]\!](\omega) = [\![\mathscr{S}]\!](\omega)$ on closed $\omega$'s? The main difficulty is that the normal bundle of $\mathscr{S}$ need not be trivial. This problem was one of the reasons for the theory of fiber bundles developed by H. Hopf and Whitney. Alternatively, can one find a $\mathscr{S}$ corresponding to a given $\sigma$? Such $\mathscr{S}$, if it exists, is by definition a current. For some $\mathscr{N}$'s and $\sigma$'s, while there exists no submanifold $\mathscr{S}$ which does the job, there will exist, however, a surface which is a submanifold except for a small singular set—this situation was one of the reasons why de Rham introduced the theory of currents in the first place, and it also was one of the reasons why Thom developed his cobordism theory.

One of the major advantages of working with currents is that a current $[\![\mathscr{S}]\!]$ always has a well defined boundary current $\partial[\![\mathscr{S}]\!]$, defined by setting $\partial[\![\mathscr{S}]\!](\varphi) = [\![\mathscr{S}]\!](d\varphi)$, i.e., Stokes's theorem becomes a definition. This notion of boundary is preserved unaer weak convergence of currents.

The *integral currents* in geometric measure theory are those currents which come from integration of forms over oriented rectifiable sets and whose current boundaries come from a similar such integration. These currents were introduced in a seminal paper, *Normal and integral currents*, by H. Federer and W. H. Fleming in 1960. Their paper **[FF]** (which was awarded one of the 1987 Steele Prizes of the American Mathematical Society) begins

"Long has been the search for a satisfactory analytic and topological formulation of the concept '$k$-dimensional domain of integration in euclidean $n$-space.' Such a notion must partake of the smoothness of differentiable manifolds and of the combinatorial structure of polyhedral chains with integer coefficients. In order to be useful for the calculus of variations, the class of all domains must have certain compactness properties. All these requirements are met by the *integral currents* studied in the paper."

Under very mild assumptions then a sequence of integral currents will have a subsequence which converges to an integral current—this is really a striking result since one could well imagine making up a sequence of integral currents whose limit might be a measure considerably smeared out (and thus not $k$-dimensional at all). The integral currents in a manifold form a chain complex whose homology groups it turns out are isomorphic with those of the singular homology theory. When one combines these various facts together with the present regularity theory, one concludes, for example, that *each integral homology class in any compact Riemannian manifold is representable by an integral current of least area*; *furthermore any such least area current is a smooth minimal submanifold except for a compact singular set of codimension at least two in the surface*. Corresponding theorems hold for currents spanning boundaries as in the oriented version of our original problem. As suggested above, cobordism theory provides topological obstructions to everywhere regularity.

Also the codimension two singularity estimate cannot be improved in general since holomorphic varieties in Kähler manifolds are area minimizing currents. Such complex varieties are but one example of large families of area minimizing "calibrated" geometries which have been (and are continuing to be) discovered. Calibration theory was surveyed by Morgan in a recent article in this MONTHLY **[M]**.

Incidentally, one of the goals of the Geometry Supercomputer Project (supported by the N.S.F. and the University of Minnesota) is the development of algorithms for the computation of optimal geometries of the types which arise in geometric theory.

As indicated above, there is really a lot more to geometric measure theory and its applications than is even suggested by least area problems. As a first step, however, minimal surfaces *à la* Morgan is hard to beat.

REFERENCES

[F] H. Federer, Geometric Measure Theory, Springer-Verlag, New York, 1969.

[FF] H. Federer and W. H. Fleming, Normal and integral currents, *Ann. of Math.*, 72 (1960) 458–520.

[M] F. Morgan, Area-minimizing surfaces, faces of Grassmannians, and calibrations, this MONTHLY, 95 (1988) 813–822.

# TELEGRAPHIC REVIEWS

*Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.*

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

| | | |
|---|---|---|
| T: Textbook | P: Professional Reading | 1–4: Semesters |
| C: Computer Software | L: Undergraduate Library | **: Special Emphasis |
| S: Supplementary Reading | 13: Grade Level | ??: Questionable |

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

**General, S*(13-14), L*.** *Learning and Doing Mathematics.* John Healey Mason. Macmillan Education (US Distr: Scholium Intern), 1988, vii + 82 pp, £5.95 (P). [ISBN: 0-333-44942-8] Text discusses specializing and generalizing as strategies for doing and learning mathematics. Good supplement for first course in which students are expected to read mathematical text and write proofs. Examples and exercises assume only high school mathematics. Based on a unit of Open University Foundations Course in Mathematics. KS

**General, S.** *Making Connections with Mathematics.* John Egsgard, *et al.* Janson, 1988, x + 102 pp, $19.95 (P). [ISBN: 0-939765-27-6] Thirty-nine "activities" designed to encourage students to think about the uses of mathematics in the real world (e.g., nature, science, space, computer science, business, music). Here are ideas for class exploration, discussion, and investigation suitable for courses at the secondary level (general mathematics, algebra, trigonometry, number theory, probability, and statistics). LCL

**General, S, L*.** *"What Do You Care What Other People Think?" Further Adventures of a Curious Character.* Richard P. Feynman. WW Norton, 1988, 255 pp, $17.95. [ISBN: 0-393-02659-0] Second volume of Feynman's autobiographical essays. Over half the volume is devoted to a characteristically blunt account of Feynman's investigative work for the Rogers Commission which reported on the Challenger explosion. One can see the mind of a scientist at work in the questions he asked—many of which are rooted in basic numeracy (e.g., estimating probabilities, measuring roundness, verifying mathematical models of temperature). LAS

**General, S**, L**.** *Mathematical Writing.* Donald E. Knuth, Tracy Larrabee, Paul M. Roberts. MAA Notes, No. 14. MAA, 1989, 115 pp, (P). [ISBN: 0-88385-063-X] A fascinating set of lecture notes from a course given at Stanford by Donald Knuth, full of good advice and clever examples. Hints cover mathematical writing, typing, typesetting, literate programming, publication, refereeing, and illustrations. Guest lectures by Leslie Lamport (author of LaTeX), Mary-Claire von Leuner, and *Monthly* editors Herb Wilf and Paul Halmos. Candid, witty, opinionated; rooted in real examples of real authors and real editors writing for real audiences. Includes mini-essays and running commentary on use of "hopefully" and "which," as well as a masterful set of composition assignments designed to exercise one's command of words. LAS

**General, P.** *Ten Papers Translated from the Russian.* M.B. Abalovich, *et al.* AMS Transl., Ser. 2, V. 142. AMS, 1989, vi + 121 pp, $52. [ISBN: 0-8218-3122-4]

**Mathematics Appreciation, S(13).** *Glyphs: Getting the Picture.* William Sacco, *et al.* Contemp. Appl. Math. Janson, 1987, vii + 49 pp, (P). [ISBN: 0-939765-07-1] This student workbook could be used in liberal arts or secondary mathematics courses to introduce glyphs, pictures used to represent data. Variety of examples, many from the medical field. Short on initial explanation, motivation. MW

**Precalculus, T(13: 1).** *Mathematics for Calculus.* James Stewart, Lothar Redlin, Saleem Watson. Brooke/Cole, 1989, xviii + 589 pp, $39.25. [ISBN: 0-534-10080-5] Basic precalculus text emphasizing problem solving. Extensive treatment of trigonometry. Includes a chapter on each of systems of equations and inequalities, analytic geometry, and sequences and series. SP

**Precalculus, T(13: 1).** *Basic Mathematics.* Serge Lang. Springer-Verlag, 1988, xv + 475 pp, $29.80 (P). [ISBN: 0-387-96787-7] Paperback reprint of a text first published by Addison-Wesley in 1971 (TR, February 1972). Concise, clearly written text for course preceding calculus. Covers algebra, intuitive and coordinate geometry, functions, complex numbers, induction, determinants on both manipulative

and theoretical levels. Exercise sets include many proofs. KS

**Precalculus, T*(13).** *Algebra and Trigonometry with Analytic Geometry, Seventh Edition.* Earl W. Swokowski. PWS-Kent, 1989, xii + 668 pp. [ISBN: 0-534-91712-7] Changes to this standard text are aimed at making it easier for the student to use. These include expanded solutions to examples, extensive revision of the logarithm section, a reworking of many aspects of trigonometric functions. In addition, approximately half the exercises differ from the last edition, including the insertion of about 140 new applied exercises. (*Fourth Edition,* TR, December 1978; *Sixth Edition,* TR, May 1987.) MR

**Education, P.** *The Van Hiele Model of Thinking in Geometry Among Adolescents.* David Fuys, Dorothy Geddes, Rosamond Tischler. J. for Res. in Math. Educ., Mono. No. 3. NCTM, 1988, x + 195 pp, $7.50 (P). Report on research carried out by the Brooklyn College Project on the van Hiele levels model of thinking. Describes the Project's research tools and student performance using these tools. Presents analysis and summary of both student use and teacher use of tools in an analysis of geometry contained in three U.S. textbook series. Ends with implications of study, in particular, for classroom practice, teacher training, and curriculum design. Bibliography. RJA

**Education, P, L*.** *Reshaping College Mathematics: A Project of the Committee on the Undergraduate Program in Mathematics.* Ed: Lynn Arthur Steen. MAA Notes No. 13. MAA, 1989, x + 124 pp, (P). [ISBN: 0-88385-062-1] A compendium of CUPM reports of the last decade, covering the full spectrum of undergraduate courses. Each report is preceded by a new preface, as is the whole volume. Eleven chapters in all: the first six reprint the out-of-print 1981 CUPM *Recommendations for a General Mathematical Science Program*; the final five reprint various other CUPM reports on discrete mathematics, mathematics appreciation courses, and transition from high school to college (three reports). From the Editor's Preface: "We don't need to look far for sound goals and objectives for college mathematics. Most of what we need can be found in this volume." LW

**Education, S(16-18).** *Imaginative Ideas for the Teacher of Mathematics, Grades K-12: Ranucci's Reservoir.* Ed: Margaret A. Farrell. NCTM, 1988, iv + 92 p, $8 (P). [ISBN: 0-87353-257-0] Collection of articles from various sources written by Ernest Ranucci. Sections on pattern finding, applications, spatial visualization and drawing, inventiveness, and games. Bibliography. Many teaching ideas. MW

**Education, S(15-18).** *Problem Solving: Tips for Teachers: Selections from the Arithmetic Teacher.* Ed: Phares G. O'Daffer. NCTM, 1988, iv + 83 pp, $7 (P). [ISBN: 0-87353-264-3] Thirty-six articles reprinted from the last five years of *Arithmetic Teacher.* Activities to develop and extend problem-solving strategies and skills in elementary students, and hints on implementing a problem-solving program. Each two or three-page article contains problems, teaching suggestions, ideas for classroom climate, and pointers to additional resources. MW

**Education, P*.** *The Use of Calculators in the Standardized Testing of Mathematics.* Ed: John W. Kenelly. MAA Notes No. 12. The College Board and MAA, 1989, vi + 50 pp, $6.50 (P). Papers and policy recommendations from a November 1986 joint symposium sponsored by MAA and the College Board to develop a consensus statement on the role of calculators in mathematics tests. Background papers by Kenelly, Wilson, Kilpatrick, Leitzel, Waits, and Harvey discuss aspects of policy, history, psychometric issues, experience with calculator-sensitive items, and implications for actual tests. Concludes with a statement of symposium recommendations and some sample tests. LAS

**Education, P, L*.** *New Directions for Elementary School Mathematics: 1989 Yearbook.* Eds: Paul R. Trafton, Albert P. Shulte. NCTM, 1989, viii + 245 pp, $18. [ISBN: 0-87353-272-4] Twenty-one papers on the process of change, on how children think, on new approaches to content, on strategies for exploration, and on special factors that influence teaching. The papers cover much the same ground as other recent studies of school mathematics, but less boldly. Many papers give good examples of active, instructional strategies; few papers probe as well the instructional impact of calculators. There is much good but little new in these *New Directions.* LAS

**Education, S, P, L*.** *Research Issues in the Learning and Teaching of Algebra.* Ed: Sigrid Wagner, Carolyn Kieran. Research Agenda for Math. Educ., V. 4. NCTM and Lawrence Erlbaum Associates, 1989, x + 287 pp, $15 (P). [ISBN: 0-87353-268-6] Twenty-three papers from a March 1987 conference at the University of Georgia inquiring into what algebra is and should become in an age of computer algebra systems; what research reveals about teaching and learning of algebra; and how algebraic thinking relates to mathematical thinking. Authors include Thorpe, Leitzel, Tall, Fey, Davis, Wheeler, and many others. LAS

**Education, T(16-18), S, L.** *Cooperative Learning in Mathematics: A Handbook for Teachers.* Ed: Neil Davidson. Addison-Wesley, 1990, 409 pp, (P). [ISBN: 0-201-23299-5] Seventeen authors offer concrete suggestions for implementing small group discussion in mathematics classrooms, elementary through college. Summaries of research results, model programs, sample activities, bibliographies, and list of resource materials. Valuable for both pre-service and in-service programs. MW

**Education, P, L.** *Mathematical Enculturation: A Cultural Perspective on Mathematics Education.* Alan J. Bishop. Math. Educ. Lib. Kluwer Academic, 1988, xiv + 195 pp, $69. [ISBN: 90-277-2646-9] An examination of the common cultural roots of mathematics in all societies—counting, locating, measuring, designing, playing, explaining—together

with implications of these commonalities for mathematics education (or "enculturation"). Bishop argues, based on this broad notion of mathematical culture, that the ideal mathematics educator would be more like an advocate than a mathematician—a person with commitment and the ability to communicate who can personify mathematical culture and be accountable to it, but who may not be part of it. LAS

**History, P, L**.** *A Century of Mathematics in America.* Eds: Peter Duren, Richard A. Askey, Uta C. Merzbach. AMS, 1989. *Part I.* History of Math., V. 1, viii + 477 pp, $57 [ISBN: 0-8218-0124-4]; *Part II.* History of Math., V. 2, x + 585 pp, $70. [ISBN: 0-8218-0130-9] An extraordinary collection of over 70 papers—some new, some reprints—covering personal reminiscences, historical surveys of fields, autobiographical essays, departmental histories, and pressures of war, politics, and emigration. Altogether a rich collage of the history of mathematics in America, mixing accounts of research, people, teaching, organizations, politics, and education, all expressed through first-hand accounts by the giants of American mathematics. LAS

**History, P.** *Die Mathematik und ihre Dozenten an der Berliner Universität, 1810-1933.* Kurt-R. Biermann. Akademie-Verlag, 1988, 391 pp, 98 DM. [ISBN: 3-05-500402-7] A detailed and very interesting history of mathematics at the University of Berlin up to 1933. Much information on appointments and promotions and how they were made. JD-B

**History, T(16: 1), L*.** *A History of Mathematics, Second Edition.* Carl B. Boyer, Uta C. Merzbach. Wiley, 1989, xviii + 762 pp, $53.95. [ISBN: 0-471-09763-2] First twenty-two chapters are essentially unchanged from Boyer's 1968 Wiley original (TR, January 1969; 1985 Princeton U. Pr. paperback, TR, February 1986). Chapters dealing with the 19th and 20th centuries have been revised and expanded. References and general bibliography have been revised. Foreign-language chapter references have been dropped. The Boyer style and approach remain. Comprehensive, interesting, and readable with appropriate exercises. JK

**Graph Theory, T(15-17: 1, 2), L.** *Graph Theory.* Ronald Gould. Benjamin/Cummings, 1988, x + 332 pp, $39.95. [ISBN: 0-8053-6030-1] An introductory course, concentrating on well-established topics (paths, searching, trees, networks, planarity, matching, independence, extremal theory, applications), with emphasis on algorithmic aspects (no programming prerequisites) as well as theoretical results. Versatile text, suitable for one or two semesters. LCL

**Number Theory, T*(13: 1), S, C, L*.** *Invitation to Number Theory with Pascal.* Donald D. Spencer. Camelot, 1989, 320 pp, $19.95 (P) [ISBN: 0-89218-126-5]; *Instructor's Manual to Accompany,* 1989, 164 pp, (P). [ISBN: 0-89218-127-3] Highly readable introduction to both Pascal and number theory. Includes discussions of primes, perfect numbers, Fibonacci numbers, $\pi$, Pythagorean triples, figurate numbers, magic squares, number systems, modular arithmetic, palindromes, Pascal's triangle, factoring, mathematical puzzles, amicable numbers, and many other interesting number types. Includes lots of exercises and selected answers. CEC

**Linear Algebra, T?, S, L.** *Hypercomplex Numbers: An Elementary Introduction to Algebras.* I.L. Kantor, A.S. Solodovnikov. Transl: A. Shenitzer. Springer-Verlag, 1989, x + 169 pp, $39.80. [ISBN: 0-387-96980-2] The first part of the book introduces algebras as generalizations of the complex numbers. A hypercomplex number system is a set of numbers of the form $a_0 + a_1 i_1 + \ldots + a_n i_n$, where $a_k$ is a real number and $i_k$ is an imaginary unit. Discusses when these systems allow division or sums of squares identities. The rest of the book covers algebras, vector spaces, systems of equations, scalar products, orthogonal transformations, and normed algebras. Terse text may be difficult for many to read. SB

**Group Theory, S(18), P.** *Endliche p-Gruppen.* László Rédei. Akademiai Kiado, 1989, 304 pp, $36. [ISBN: 963-05-4660-4] A monograph based quite largely on the author's own research. JD-B

**Group Theory, P.** *Abelian Group Theory.* Eds: Laszlo Fuchs, Rüdiger Göbel, Phillip Schultz. Contemp. Math., V. 87. AMS, 1989, xvii + 289 pp, $33 (P). [ISBN: 0-8218-5068-7] A collection of twenty-six papers delivered at a 1987 conference in Perth which captures the current state of abelian group theory. Included is an essay by R. Göbel on the work of Ulm and its impact on abelian group theory. SG

**Algebra, P.** *Invariant Theory.* Ed: R. Fossum, *et al.* Contemp. Math., V. 88. AMS, 1989, ix + 598 pp, $52 (P). [ISBN: 0-8218-5094-6] A collection of over twenty papers presented at an AMS Special Session in 1986 focusing on algebraic group actions on algebraic varieties. SG

**Algebra, P.** *Nombre de Points d'une Courbe sur un Corps Fini.* Jean-Pierre Soublin. Papers in Pure & Appl. Math., No. 79. Queen's U, 1988, 158 pp, (P). Suppose $E$ is a polynomial in two variables with integer coefficients. The author studies methods for finding good estimates for the number of solutions of $E = 0$ over fields of prime order. Although this volume contains no new results, it explains previous results in simpler terms. CEC

**Calculus, T(13).** *Elements of Calculus, Second Edition.* G. Don Allen, Charles Chui, Bill Perry. Brooks/Cole, 1989, x + 534 pp, $38. [ISBN: 0-534-09240-3] Changes from the *First Edition* (TR, October 1983) include better problem sets, e.g., more even-odd pairing of like exercises and fewer problems involving tedious calculations; additional chapters on infinite series and probability; and sections on computer applications. As with the *First Edition,* this text is aimed at students of business, social sciences, and biology. MR

**Calculus, T(14-15: 2), P.** *Advanced Calculus for Users.* Alain Robert. North-Holland (US Distr: Elsevier Science), 1989, xvii + 364 pp, $60.50. [ISBN: 0-444-87324-4] Applications-oriented. Written for

physicists, engineers, and others who use results of analysis in their work. Main topics are differentiability, integration of differential forms, function spaces, and Fourier series. Emphasis on linearization. Appendix includes convergence results as well as a number of exercises. SB

**Calculus, T(13-14: 2).** *Applied Calculus.* Bernard Kolman, Charles G. Denlinger. Harcourt Brace Jovanovich, 1989, xiv + 895 pp, $33. [ISBN: 0-15-502903-7] Calculus for students majoring in business, economics, and the management, life, and social sciences. Informal, intuitive, geometric; theory is minimized. Contains a multitude of applied examples and exercises drawn from the targeted majors. An optional Appendix reviews basic algebra. Extends the authors' *Calculus for the Management, Life, and Social Sciences, Second Edition* (TR, January 1989) by adding double integrals, differential equations, probability theory, Taylor polynomials, and infinite series. DFA

**Real Analysis, T(15: 2), L.** *Foundations of Analysis: The Theory of Limits.* Herbert S. Gaskill, P.P. Narayanaswami. Harper & Row, 1989, xv + 639 pp. [ISBN: 0-06-044734-6] Text for first rigorous real analysis course. Covers the essential elements of analysis of the real line by focusing on the epsilon-delta definition of limit. Final two chapters extend these ideas into the plane and metric spaces. The discussion of ideas is extensive and stresses the role of algebra. Nearly 3000 exercises. SP

**Real Analysis, T(14-15: 1), L.** *An Introduction to Analysis.* James R. Kirkwood. PWS-Kent, 1989, xii + 276 pp. [ISBN: 0-534-91500-0] Introduction to theory of real-valued functions in one variable intended in particular for mathematics students who are making the transition from computational to theoretical courses. Numerous exercises ranging from routine to difficult with hints and solutions to many of them provided in the back. LC

**Real Analysis, T(16: 1), S, L.** *Intermediate Mathematical Analysis.* Hugh Thurston. Clarendon Pr, 1988, ix + 164 pp, $55; $24.95 (P). [ISBN: 0-19-853291-1; 0-19-853292-X] A rigorous introduction to multivariable calculus which assumes some background in elementary real analysis. Includes differentiation, Taylor's theorem, Lagrange multipliers, the implicit function theorem, the function theorem, and integration. Includes exercises along with hints and solutions. CEC

**Complex Analysis, S(18), P.** *Cauchy-Riemann (CR) Manifolds.* Geraldine Taiani. Pace U (Math. Dept., NY City Campus, NY 10038), 1989, 102 pp, $15 (P). *CR* geometry is a complex analog of Riemannian geometry, applied, e.g., to boundary extension problems for functions of several complex variables. Although papers in the subject have appeared for a decade or so, basic definitions and theorems have not before now been collected in a single source. This informal (and informally printed) monograph, in five short chapters and a bibliography, fills that gap. A useful introduction to an active research area. PZ

**Differential Equations, S(17), P.** *Continuous Decoupling Transformations for Linear Boundary Value Problems.* P.M. van Loon. CWI Tract, V. 52. Mathematisch Centrum, 1988, vi + 198 pp, Dfl. 30.40 (P). [ISBN: 90-6196-353-2] If a boundary value problem has both (fast) increasing and (fast) decaying modes, integration can be expensive and/or inaccurate. The problem can be avoided by decoupling these modes using transform-like techniques and then working with the modes separately. However, new problems such as non-linearity arise. The technique presented here, which the author claims is robust, uses a combination of a Riccati transform with an invariant imbedding. MR

**Differential Equations, S(14-15).** *Schaum's Solved Problems Series: 2500 Solved Problems in Differential Equations.* Richard Bronson. McGraw-Hill, 1988, v + 440 pp, $19.95 (P). [ISBN: 0-07-007979-X] Problems from any area normally encountered in a first course in differential equations. Minimal discussion of theory, but an excellent source of problems for both students and instructors. Large number of applications. MR

**Differential Equations, S*, C, L*.** *Differential and Difference Equations through Computer Experiments, Second Edition.* IBM PC. Hüseyin Koçak. Springer-Verlag, 1989, xvii + 224 pp, $49.95 (P). [ISBN: 0-387-96918-7] A re-issue of the software package Phaser in versions that use the higher resolution IBM EGA/VGA graphics boards and with options to use a mathematics coprocessor. The software package and the accompanying guide book is otherwise unchanged from the 1986 *First Edition* (TR, April 1987). Phaser simulates dynamical systems by plotting solutions to differential and difference equations. It runs only on PC compatible equipment, but utilizes effective window environments. A powerful yet inexpensive exploratory tool for students, teachers, scientists, and researchers. LAS

**Differential Equations, T**(15-16), L.** *Ordinary Differential Equations, Fourth Edition.* Garrett Birkhoff, Gian-Carlo Rota. Wiley, 1989, xi + 399 pp, $67.20. [ISBN: 0-471-86003-4] The only changes in this wonderful text are the introduction of numerical methods in the first chapter and the addition of detailed reviews of elementary facts for reference. This makes an excellent text for a second course in differential equations. (*Third Edition,* TR, April 1979.) MR

**Partial Differential Equations, S(18), P.** *Solution of Initial Value Problems in Classes of Generalized Analytic Functions.* Wolfgang Tutschke. Springer-Verlag, 1989, 188 pp, $49.80 (P). [ISBN: 0-387-50216-5] The classical Cauchy-Kovalevska theorem gives sufficient conditions for solvability of certain boundary value problems in partial differential equations with holomorphic initial data. Here the author's object is to generalize the C-K theorem to problems with initial functions of generalized analytic type. The main tools are functional-analytic,

including scales of Banach spaces. Reasonably self-contained: includes, e.g., some rudiments of complex analysis in several variables. PZ

**Partial Differential Equations, T(18), P.** *Initial-Boundary Value Problems and the Navier-Stokes Equations.* Heinz-Otto Kreiss, Jens Lorenz. Pure & Appl. Math., V. 136. Academic Pr, 1989, xi + 402 pp, $54.50. [ISBN: 0-12-426125-6] Develops a theory for linear and nonlinear equations for problems well-suited to computation, examining smoothness and admissible boundary conditions in great detail. Illustrates the theory with careful study of compressible and incompressible Navier-Stokes equations. DFA

**Partial Differential Equations, S(18), P.** *Parabolic Equations on an Infinite Strip.* N.A. Watson. Pure & Appl. Math., V. 127. Marcel Dekker, 1989, xii + 289 pp, $99.75. [ISBN: 0-8247-7999-1] A study of second order, linear, parabolic, partial differential equations on an infinite strip. Emphasis on integral representations. Topics include the semigroup property, the Cauchy problem, and limit considerations. SP

**Partial Differential Equations, P.** *Lecture Notes in Control and Information Sciences-114: Control of Partial Differential Equations.* Ed: A. Bermúdez. Springer-Verlag, 1989, ix + 318 pp, $50.90 (P). [ISBN: 0-387-50495-8] Proceedings of the IFIP WG 7.2 Working Conference held at Santiago de Compostela, Spain, July 1987. MR

**Partial Differential Equations, P.** *Nonlinear Evolution Equations and Applications.* Gheorghe Moroşanu. Transl: Gheorghe Moroşanu. Math. & Its Applic. D Reidel (US Distr: Kluwer Academic), 1988, xii + 340 pp, $96. [ISBN: 90-277-2486-5] A systematic exposition of the stability considerations for nonlinear, monotone evolution equations. The idea is to carry over the stability theory from ordinary differential equations where a vector quantity changes with time to the case where a quantity in function space varies, the time rate of change being given by a monotone, nonlinear operator applied to the quantity. Discrete cases and many applications also considered. SP

**Partial Differential Equations, T(17-18: 2).** *Applied Partial Differential Equations.* Paul Du Chateau, David Zachmann. Harper & Row, 1989, xiv + 620 pp. [ISBN: 0-06-041772-2] Introductory text reflecting modern approaches to this classical subject. Approximately one-half of the book is devoted to numerical methods. Includes treatment of Fourier series, integral transforms, and finite difference methods. Applications to flow in a porous medium, wave propagation, advection and traffic flow, in addition to the more traditional heat equation and vibrating strings. SP

**Numerical Analysis, T(14-15: 1), L.** *Numerical Methods for Engineers and Computer Scientists.* Paul F. Hultquist. Benjamin/Cummings, 1988, xiii + 326 pp, $39.95. [ISBN: 0-8053-4652-X] Seeks to create an understanding of algorithms without belaboring the fine points of their construction. More of a book on numerical methods than numerical analysis. Includes Pascal code for all the algorithms. A diskette is also available. Topics covered are standard for an undergraduate numerical analysis text. SM

**Numerical Analysis, T(15-17: 1).** *Numerical Analysis for Applied Mathematics, Science, and Engineering.* Donald Greenspan, Vincenzo Casulli. Addison-Wesley, 1988, ix + 341 pp, $43.25. [ISBN: 0-201-09286-7] Revision of usual topics, to stress those needed by scientists. Balance of theory and practice, method-oriented with illustrative examples, doable exercises. Treats algebraic systems, initial value problems and boundary value problems for ordinary differential equations, elliptic, parabolic, and hyperbolic equations, nonlinear equations, and Navier-Stokes equations. RM

**Numerical Analysis, T(13-14: 1), L.** *Numerical Analysis, Algorithms and Computation.* J. Murphy, D. Ridout, Bridgid McShane. Math. & Its Applic. Halsted Pr, 1988, viii + 148 pp, $49.95. [ISBN: 0-470-21214-4] Introduction to numerical analysis; prerequisite is calculus, although knowledge of linear algebra is helpful. Covers iterative methods for nonlinear equations, interpolation, numerical integration, numerical solutions of ordinary differential equations, and systems of linear equations. LC

**Numerical Analysis, P, L.** *C Tools for Scientists and Engineers.* Louis Baker. McGraw-Hill, 1989, xii + 324 pp, $29.95 (P). [ISBN: 0-07-003355-2] A collection of C programs implementing a variety of standard numerical techniques: *LU* factorization, eigenvalue and discriminant analysis, singular value decomposition, Newton-Raphson methods for nonlinear equations, *B*-spline interpolation, adaptive quadrature, fast Fourier transforms, and methods for solving systems of ordinary differential equations. Also discusses how to convert Fortran programs into C. AO

**Numerical Analysis, T*(15: 1), L.** *Introduction to Pascal for Computational Mathematics.* E.J. Redfern. Macmillan Education (US Distr: Scholium Intern), 1987, x + 290 pp, $26.50 (P). [ISBN: 0-333-44431-0] Not only a complete introduction to Pascal, but also a nice collection of applications of programming to mathematics. For the new math major. The examples and exercises illustrate numerical methods for studying sequences and series, nonlinear algebraic equations, integration, simultaneous linear equations, first-order differential equations, elementary statistics; and computational techniques for polynomial manipulation, queues, trees, graphs. A book designed for a course which could be very interesting. DFA

**Numerical Analysis, T(15: 1).** *An Introduction to Numerical Computations, Second Edition.* Sidney Yakowitz, Ferenc Szidarovszky. Macmillan, 1989, xiv + 462 pp, $35. [ISBN: 0-02-430821-8] From a course for engineering students, but of wider interest. Simultaneous linear equations, interpolation, numerical differentiation and integration, nonlinear

equations, function approximation, differential equations. Provides solid foundation for further study, omits rigorous development of theoretical structure and error analysis. Assumes calculus and a Fortran or Pascal course. Careful discussion of round-off errors. Subroutine and program listings throughout. Lots of problems from many applied areas. DFA

**Operator Theory, T(17-18: 1), S, P.** *Nonlinear Equations and Operator Algebras.* Vladimir A. Marchenko. Transl: V.I. Rublinetsky. Math. & Its Applic. D Reidel (US Distr: Kluwer Academic), 1988, xvi + 157 pp, $64. [ISBN: 90-277-2654-X] Presents the author's method for integrating nonlinear equations (e.g., KdV, KP, NLS, etc.) by jacking them into an operator-valued setting, solving there, and projecting into special finite-dimensional spaces. BC

**Functional Analysis, T(17-18: 1), S, P, L.** *Topics in Functional Analysis and Applications.* S. Kesavan. Wiley, 1989, xii + 267 pp, $24.95. [ISBN: 0-470-21050-8] Develops basic tools of functional analysis for applications to partial differential equations via distribution theory. Applications to evolution equations and weak solutions of elliptic boundary value problems. Includes chapter-ending comments and exercises. BC

**Functional Analysis, S(18), P.** *Direct Methods in the Calculus of Variations.* Bernard Dacorogna. Appl. Math. Sci., V. 78. Springer-Verlag, 1989, ix + 308 pp, $59. [ISBN: 0-387-50491-5] Concise, largely self-contained monograph on theory of direct methods having work of Hilbert, Lebesgue, and Tonelli as historical antecedents. Includes more recent topics such as vector-valued functions and relaxation of non-convex problems. States without proof many prerequisite theorems from functional analysis. GG

**Functional Analysis, P.** *Space Mappings with Bounded Distortion.* Yu. G. Reshetnyak. Trans. of Math. Mono., V. 73. AMS, 1989, xv + 362 pp, $129. [ISBN: 0-8218-4526-8] A $n$-dimensional map is of bounded distortion if, along with regularity conditions, there is a constant $q, 1 \leq q < \infty$, such that each infinitesimally small sphere is transformed into either a point or an ellipsoid for which the ratio of largest semiaxis to the smallest is less than $q$. This book studies these functions using the fact that the components satisfy certain elliptic partial differential equations. MR

**Analysis, P.** *Lecture Notes in Mathematics-1343: Fixed Point Theory of Parametrized Equivariant Maps.* Hanno Ulrich. Springer-Verlag, 1988, 147 pp, $16.30 (P). [ISBN: 0-387-50187-8] Develops an algebraic fixed point theory for maps with symmetry properties. BC

**Analysis, T*(16-18: 1, 2), S, L.** *Fourier Analysis.* James S. Walker. Oxford U Pr, 1988, xix + 440 pp, $49.95. [ISBN: 0-19-504300-6] This textbook combines the basic mathematical theory of Fourier analysis with concise discussions of its applications (e.g., vibrations and sound, heat conduction, optics, CAT scanning). This is a very attractive book: pleasing format on glossy pages together with an inviting and relaxed presentation. LCL

**Analysis, T(17-18: 1), S, P, L.** *Plateau's Problem and the Calculus of Variations.* Michael Struwe. Math. Notes, V. 35. Princeton U Pr, 1989, x + 148 pp, $19.50 (P). [ISBN: 0-691-08510-2] Plateau's problem—to find a "minimal" surface spanning a Jordan curve in $R^n$—and the constant-mean-curvature problem in $R^3$ are considered via the calculus of variations. Unstable parametric solutions are featured. BC

**Analysis, T, S*( 13-16), P, L**.** *Polynomials.* E.J. Barbeau. Problem Books in Math. Springer-Verlag, 1989, xxii + 441 pp, $59. [ISBN: 0-387-96919-5] Beautifully composed exercise sets, proceeding from the concrete to the general, designed to entice the student into the discovery of the basic ideas in the theory of equations (evaluation and factorization of polynomials, solutions of equations, approximation and locations of zeros, symmetric functions of the zeros). In addition, 69 "Explorations" raise open-ended questions that encourage investigation, plus more that 300 "Problems" drawn from problem sections of various undergraduate journals and related contest literature. Included are hints and solutions to selected exercises and solutions to all problems. LCL

**Analysis, S(15-18), P, L.** *Means and Their Inequalities.* P.S. Bullen, D.S. Mitrinović, P.M. Vasić. Math. & Its Applic. D Reidel (US Distr: Kluwer Academic), 1988, xix + 459 pp, $89. [ISBN: 90-277-2629-9] An encyclopedic treatment of means that occur in the theory of inequalities, including historical connections and proofs of basic results. The practice of obtaining good estimates using inequalities is an art that tends to resist classification and systematization, so that this volume, which creates some order in this regard, is a welcome addition to the literature. A useful reference for scientists, mathematicians, and problem-solvers and all those who use inequalities and estimates. LCL

**Analysis, P.** *Lecture Notes in Mathematics-1359: Harmonic Analysis.* Ed: P. Eymard, J.-P. Pier. Springer-Verlag, 1988, viii + 287 pp, $24.30 (P). [ISBN: 0-387-50524-5] 26 papers on aspects of harmonic analysis form the proceedings of an international symposium held at the Centre Universitaire de Luxembourg in September 1987. A paper of J.-P. Pier gives a short history of the evolution of the subject, with an extensive bibliography; a long research-expository paper of G. Mackey treats a variety of applications. PZ

**Analysis, L.** *Nonstandard Analysis and its Applications.* Ed: Nigel Cutland. London Math. Soc. Stud. Texts, V. 10. Cambridge U Pr, 1988, xiii + 346 pp, $19.95 (P). [ISBN: 0-521-35947-3] A collection of ten papers from a conference held at the University of Hull in 1986. Several of the papers are introductory while others present a sample of recent research results. AO

**Algebraic Geometry, P.** *Algebraic Geometry and*

*Commutative Algebra in Honor of Masayoshi Nagata.* Ed: Hiroaki Hijikata, *et al.* Academic Pr, 1988, $69 each. *Volume I,* 404 pp [ISBN: 0-12-348031-0]; *Volume II,* 399 pp. [ISBN: 0-12-348032-0] Papers dedicated to Nagata on the occasion of his sixtieth birthday. The papers cover a wide range of topics and are of great interest to anyone in either of the two fields. SG

**Algebraic Geometry, T(16-18: 1), S, P, L.** *Geometry and Codes.* V.D. Goppa. Math. & Its Applic. Kluwer Academic, 1988, ix + 157 pp, $64. [ISBN: 90-277-2776-7] Algebraic geometry and the Riemann-Roch theorem take on the task of constructing and analyzing error-correcting codes. A nicely balanced introduction to both subjects, with lots of explicit examples of curves and codes. BC

**Differential Geometry, P.** *Lecture Notes in Mathematics-1368: Traces of Differential Forms and Hochschild Homology.* Reinhold Hübl. Springer-Verlag, 1989, 111 pp, $14.30 (P). [ISBN: 0-387-50985-2] Begins with the study of the Hochschild homology functor from commutative algebras to anticommutative *DG*-algebras. Results presented include the existence of functional homomorphisms in various settings, construction of a trace for the Hochschild homology of topological pairs, and a proof that Lipman's "trace formula II" holds for local complete intersections. MR

**Differential Geometry, S(18), P.** *The Geometry of Jet Bundles.* D.J. Saunders. London Math. Soc. Lect. Note Ser., V. 142. Cambridge U Pr, 1989, 293 pp, $29.95 (P). [ISBN: 0-521-36948-7] An introduction to jet bundles, generalizations of tangent vectors, and bundles for mathematicians and physicists interested in differential equations. Assumes differential geometry, but reviews bundles. Includes index of notation. GG

**Geometry, T*.** *Introduction to Geometry, Second Edition.* H.S.M. Coxeter. Classics Library. Wiley, 1989, xvii + 469 pp, $34.50 (P). [ISBN: 0-471-50458-0] A paperback reprint of the 1969 classic (Wiley hardcover edition, TR, January 1970). JNC

**Geometry, S(16-17), P, L.** *Mechanical Geometry Theorem Proving.* Shang-Ching Chou. Math. & Its Applic. D Reidel (US Distr: Kluwer Academic), 1988, xii + 302 pp, $79. [ISBN: 90-277-2650-7] From the Foreword by Larry Wos: "Although in no way should one assume that the problem of proving theorems in geometry is completely solved, Chou has demonstrated that a very large number of theorems can be proved by a single program relying on a well-defined paradigm." The paradigm is algebraic geometry and the Gröbner basis method. This must be what Descartes had in mind. BC

**Geometry, T(15-16: 1), S, L.** *Tilings and Patterns: An Introduction.* Branko Grünbaum, G.C. Shephard. WH Freeman, 1989, ix + 446 pp, $27.95 (P). [ISBN: 0-7167-1998-3] Paperback reprint of the first seven chapters of the 1987 twelve-chapter *Tilings and Patterns* (TR, August-September 1987; Extended Review, January 1988). Covers topological and symmetry properties of tilings; tilings by regular polygons; motif-transitive patterns; and several classifications of very symmetric tilings. Includes full bibliography of original volume. The definitive resource for information about patterns in the plane. LAS

**Algebraic Topology, S(18), P.** *Lecture Notes in Mathematics-1361: Algebraic Topology and Transformation Groups.* Ed: T. tom Dieck. Springer-Verlag, 1988, vi + 298 pp, $28.60 (P). [ISBN: 0-387-50528-8] Proceedings of a conference held in Göttingen in August 1987. Eleven papers; list of participants. JS

**Topology, T(18: 1), S, P.** *Infinite-Dimensional Topology: Prerequisites and Introduction.* J. van Mill. Math. Lib., V. 43. North-Holland (US Distr: Elsevier Science), 1989, xii + 401 pp, $73.25. [ISBN: 0-444-87133] Chapters 1-5 are intended to be suitable as a graduate text serving as a brief introduction to dimension theory and ANR theory, assuming a previous course in basic topology. Chapter 6 is an introduction to infinite-dimensional topology with Chapters 7-8 dealing with cell-like maps, Q-manifolds, and applications. Exercises, notes and comments, bibliography, index. JS

**Optimization, P.** *Numerical Solution of Optimal Control Problems with State Constraints by Sequential Quadratic Programming in Function Space.* K.C.P. Machielsen. CWI Tract, V. 53. Mathematisch Centrum, 1988, 214 pp, Dfl. 33 (P). Develops method analogous to sequential quadratic programming, using an iterative descent method, with search direction determined by solution of subproblems with quadratic objective function and linear constraints, penalty function determining step size at each iteration. Numerical implementation by an indirect method, solution of linear multipoint boundary value problem. Abstract treatment, solutions of some sample problems. RM

**Dynamical Systems, S*(15-16), P, L*.** *Dynamics—The Geometry of Behavior, Part 4: Bifurcation Behavior.* Ralph H. Abraham, Christopher D. Shaw. Visual Math. Lib., Vismath V. 4. Aerial Pr, 1988, xi + 196 pp, $38 (P). [ISBN: 0-942344-04-9] A picture book of dynamical systems—the concluding volume of this innovative series from the Visual Mathematics Project at the University of California at Santa Cruz: four-color hand-drawn illustrations (to emulate mathematicians' working habits) of bifurcation processes, with spare, historically-rooted captions explaining the process without using any mathematical symbols. Begins with history of investigations into the shape of the earth—whence "bifurcation"—and concludes with a pictorial atlas of catastrophes, fractals, chaos, and turbulence. A superb source of insight and motivation. (*Parts 1-3,* TR, August-September 1983; March 1984; February 1986.) LAS

**Dynamical Systems, T, S**(12-15), P, L*.** *Chaos, Fractals, and Dynamics: Computer Experiments in Mathematics.* Robert L. Devaney. Addison-Wesley, 1990, x + 181 pp, $22 (P). [ISBN: 0-201-23288-X] High-school-level introduc-

tion to quadratic iteration and the Julia and Mandelbrot sets using simple calculator and Basic programs for common school computers. Introduces by example basic ideas of dynamical systems (orbits, attractors, bifurcation, chaos), iteration in the complex plane, geometric iteration, and common fractals. Uses primarily standard topics of second-year high school algebra; operations on complex numbers are introduced as needed. A rich, challenging supplement for high-school projects, math clubs, or senior enrichment courses; would also be superb for freshman seminars in college. LAS

**Control Theory, P.** *Lecture Notes in Control and Information Sciences-115: Geometric Theory for Infinite Dimensional Systems.* H.J. Zwart. Springer-Verlag, 1989, viii + 156 pp, $26.30 (P). [ISBN: 0-387-50512-1] A monograph presenting a geometric approach to disturbance decoupling problems for infinite dimensional systems. Topics include system invariance concepts, discrete spectral systems, measurement feedback, and stability. SP

**Control Theory, P.** *Adaptive Markov Control Processes.* O. Hernández-Lerma. Appl. Math. Sci., V. 79. Springer-Verlag, 1989, xiv + 148 pp, $39.80. [ISBN: 0-387-96966-7] At each decision time the controller must estimate true parameter values and adapt the control actions to the estimated values. Recent theoretical developments; reflects author's research interests. TH

**Probability, P.** *Percolation.* Geoffrey Grimmett. Springer-Verlag, 1989, xi + 296 pp, $49.80. [ISBN: 0-387-96843-1] Percolation is the study of how processes move along paths. The mathematical formulation used here has the paths as the edges of the cubic lattice $Z^d$ where the edges are open (allowing passage) with probability $p$. The central question is: Does a particular vertex have an infinitely large connected subgraph of open edges about it? In two dimensions, the answer is yes if $p \geq 1/2$. Accessible to nonspecialists and graduate students. MR

**Probability, S(16-18), P.** *Die Entwicklung der Wahrscheinlichkeitstheorie von den Anfängen bis 1933: Einführungen und Texte.* Ivo Schneider. Akademie-Verlag, 1989, xiii + 529 pp, 79 DM. [ISBN: 3-05-500403-5] The text, in German except when first published in English, of seventy-two excerpts from the works of authors from Pseudo-Ovidius to Kolmogoroff. Short biographies. JD-B

**Stochastic Processes, P.** *Some Topics in Probability and Analysis.* Richard F. Gundy. CBMS Reg. Conf. Ser. in Math., No. 70. AMS, 1989, v + 49 pp, $15 (P). [ISBN: 0-8218-0721-8] Lecture notes from a conference at DePaul University, July 1986. Local time theory for Brownian motion, Riesz transforms, and Riesz inequalities for the infinite-dimensional version of the Ornstein-Uhlenbeck semigroup. TH

**Stochastic Processes, P.** *Equilibrium Distributions of Branching Processes.* A. Liemant, K. Matthes, A. Wakolbinger. Math. & Its Applic. Kluwer Academic, 1988, 240 pp, $69. [ISBN: 90-277-27740] Starts from theories of spatially homogeneous branching models and of spatially inhomogeneous substochastic translations. TH

**Elementary Statistics, T*(13), L.** *Elementary Statistics, Fourth Edition.* Mario F. Triola. Benjamin/Cummings, 1989, xvi + 784 pp, $35.95. [ISBN: 0-8053-0271-9] More emphasis on real data in examples and exercises. New sections cover odds and multiple regression. Chapters now include an overview, review, computer project, and data project. Custom software available, or Minitab exercises. Intended for non-math majors. Readable, includes many stories of real applications of statistics. (*Second Edition,* TR, June-July 1983; *Third Edition,* TR, June-July 1987.) TH

**Elementary Statistics, S*(13).** *Minitab Statistical Software: Student Laboratory Manual and Workbook to Accompany Elementary Statistics, Fourth Edition.* Mario F. Triola. Benjamin/Cummings, 1989, v + 164 pp, $9.95 (P). [ISBN: 0-8053-0278-6] Companion guide to Triola's text (see TR above). Illustrates appropriate Minitab procedures for each chapter and then provides "experiments" for the student to carry out. RSK

**Elementary Statistics, T?(13: 1), S.** *Interpreting Data: A First Course in Statistics.* Alan J.B. Anderson. Chapman & Hall, 1989, xvi + 223 pp, $24 (P); $55. [ISBN: 0-412-29570-9; 0-412-29560-1] Nonstandard introductory text. Probability and hypothesis testing are covered in one concise chapter. Emphasis is on collecting, summarizing, and interpretating data. RSK

**Elementary Statistics, S(13-15), L*.** *News & Numbers.* Victor Cohn. Iowa State U Pr, 1989, xii + 178 pp, $17.95; $9.95 (P). [ISBN: 0-8138-1442-1] A journalist's guide to statistics ("the science of state"), especially those that beset the health, environment, and science beats. Cohn, an award-winning *Washington Post* science writer, explains key statistical terms with concrete illustrations from real stories. He also gives numerous examples of questions for journalists to ask (How much is certain? Have the results been independently reproduced? Who paid you? Would you eat it?), as well as examples of politically-motivated changes in the definition of certain government indices. Superb supplementary reading for courses in elementary statistics. LAS

**Elementary Statistics, T(13: 1).** *Basic Statistics for Business and Economics, Second Edition.* Leonard J. Kazmier, Norval F. Pohl. McGraw-Hill, 1984, xvi + 592 pp, $29.95. [ISBN: 0-07-033448-X] Introductory statistics for students of business and economics. Answers to selected exercises in the back section. Many "real world applications" are included. Computer outputs of Minitab, SAS, and SPSS are shown in several chapters. New sections on surveys and experiments, joint probability tables, counting methods, the hypergeometric distribution, and residual plots. Marginal notes to highlight concepts have been added. (*First Edition,* TR, August-September 1980.) MS

**Elementary Statistics, T(13: 1).** *A First Course in Probability and Statistics with Applications, Second Edition.* Peggy Tang Strait. Harcourt Brace Jovanovich, 1989, xxii + 599 pp, $32. [ISBN: 0-15-527523-2] One-semester course for students with some background in calculus. Very clear explanations of concepts followed by illustrative examples and exercises. Monte Carlo methods and decision problems are included. Thorough discussion of implications of law of large numbers and central limit theorem. Answers to most exercises appear at end of book. (*First Edition,* TR, April 1984.) MS

**Computational Statistics, P*.** *Nonlinear $L_p$-Norm Estimation.* René Gonin, Arthur H. Money. Statistics: Textbooks & Mono., V. 100. Marcel Dekker, 1989, viii + 300 pp, $99.75. [ISBN: 0-8247-8125-2] Treats both the numerical and statistical aspects of the nonlinear $L_p$-norm estimation problem, particularly $L_1$-norm and $L_\infty$-norm. Begins with a concise survey of linear $L_p$-norm estimation, and concludes with three applications of nonlinear $L_p$-norm estimation. Extensive bibliography. Note price! RSK

**Computational Statistics, P, L.** *STATAL: Statistical Procedures in Algol 60.* Eds: R. van der Horst, R.D. Gill. Mathematisch Centrum, 1988. *Part 1,* CWI Syllabus, V. 20, xvii + 200 pp, Dfl. 33 (P) [ISBN: 90-6196-358-3]; *Part 2,* 197 pp, Dfl. 30.40 (P) [ISBN: 90-6196-359-1]; *Part 3,* 227 pp, Dfl. 35.40 (P). [ISBN: 90-6196-360-5] Full text of the STATAL library of statistical procedures written in Algol-60 for use on CDC Cyber computers. Over 170 procedures for distributions, computation of statistics, sorting and ranking, permutations and combinations, random number generators, tables and pictures. Each procedure is described in detail, with sample output, so would be easy to translate into other computer languages. Code numbers assigned to procedures are used both to index the library and to call the procedure from within an Algol program. LAS

**Computational Statistics, P.** *Continued Fractions in Statistical Applications.* K.O. Bowman, L.R. Shenton. Statistics: Textbooks & Mono., V. 103. Marcel Dekker, 1989, x + 330 pp, $89.75. [ISBN: 0-8247-8120-1] Gives "an account of the use of continued fractions and Padé sequences as tools in the interpretation of divergent or slowly convergent series occurring in theoretical statistics." Note price. RSK

**Statistics, T(15-17: 1, 2).** *Data Analysis for Research Designs: Analysis of Variance and Multiple Regression/Correlation Approaches.* Geoffrey Keppel, Sheldon Zedeck. WH Freeman, 1989, xxiv + 594 pp, $42.95. [ISBN: 0-7167-1991-6] Designed for behavioral and social science students. Parallel presentation of elementary ANOVA and multiple regression/correlation techniques, showing their equivalence in designed experiments. Emphasizes single-degree-of-freedom comparisons. Minimal set of exercises. RSK

**Statistics, P*.** *Statistical Methods in Accelerated Life Testing.* Reinhard Viertl. Appl. Stat. & Econom., V. 32. Vandenhoeck & Ruprecht, 1988, viii + 134 pp, DM 45 (P). [ISBN: 3-525-11266-1] Survey of statistical methods (Bayesian, as well as classical parametric and nonparametric) for determining lifetimes under normal conditions while testing under conditions of increased stress. Extensive set of references. RSK

**Statistics, P.** *Lecture Notes in Statistics-52: The Matching Methodology: Some Statistical Properties.* Prem K. Goel, T. Ramalingam. Springer-Verlag, 1989, viii + 152 pp, $20.60 (P). [ISBN: 0-387-96970-5] Monograph dealing with theoretical properties and empirical evaluations of the quality of files obtained by various procedures for merging two incomplete micro-data files, either on the same individuals or on similar individuals. RSK

**Statistics, T(17: 1), S, P.** *Lecture Notes in Statistics-50: Parametric Statistical Models and Likelihood.* Ole E. Barndorff-Nielsen. Springer-Verlag, 1988, vii + 276 pp, $28 (P). [ISBN: 0-387-96928-4] Mathematically advanced aspects and parametric statistical inference. Topics included are likelihood, transformation and exponential models, reparametrizations and differential geometry, cumulants, Laplace's method, Edgeworth approximations, and saddle-point approximations. Extensive list of references. MS

**Statistics, T(18: 2), P*.** *Nonlinear Regression.* G.A.F. Seber, C.J. Wild. Ser. in Prob. & Math. Stat. Wiley, 1989, xx + 768 pp, $59.95. [ISBN: 0-471-61760-1] Extensive, wide-ranging, predominantly theoretical treatment of nonlinear model fitting. Deals with such topics as estimation techniques and problems associated with them, curvature, autocorrelated errors, growth, compartmental, and multiphase models, errors-in-variables models, and multiresponse models. Also includes several chapters on recent algorithms for optimization and least squares. Good set of references. No exercises. RSK

**Statistics, P.** *Bayesian Analysis in Econometrics and Statistics: Essays in Honor of Harold Jeffreys.* Ed: Arnold Zellner. Robert E Krieger, 1989, xi + 474 pp, $55. [ISBN: 0-89464-354-1] Twenty-eight essays, many from the semi-annual NBER-NSF Seminar on Bayesian Inference in Econometrics and Statistics. Topics include prior distributions, adversary preposterior analysis, economic theory, applications, time series, interdependent econometric models, inference, and computer programs. TH

**Computer Literacy, T*(13-14: 1), S, L*.** *The Turing Omnibus: 61 Excursions in Computer Science.* A.K. Dewdney. Computer Science Pr, 1989, xiv + 415 pp, $24.95. [ISBN: 0-7167-8154-9] A "sneak preview" of computer science in the form of brief introductory vignettes, all mixed up in order ("not an unfair impression of computer science as a whole"), on 61 distinct topics such as algorithms, splines, fast Fourier transform, text compression, and relational databases. Rather like a selective lexicon of computer science with fascinating short es-

says about each word or phrase. Each excursion has a few exercises and references, making it possible to use the book as the text for a course in computer literacy. LAS

**Elementary Computer Science, T*(14: 1), L.** *Data Structures with Abstract Data Types and Pascal, Second Edition.* Daniel F. Stubbs, Neil W. Webre. Brooks/Cole, 1989, xxi + 471 pp, $46. [ISBN: 0-534-09264-0] For the data structures or CS 2 course: arrays, records, stacks, queues, linear structures, trees, internal sorting, sets, strings, graphs. Strong emphasis on abstract data-types. Improves specification and implementation modules of the *First Edition;* adds applications. Attractive, well-illustrated, nicely-written. DFA

**Programming, T?(13-14: 1).** *Program Design with Pseudocode, Third Edition.* Therold E. Bailey, Kris Lundgaard. Ser. in Prog. Brooks/Cole, 1989, xiii + 206 pp, $19.75 (P). [ISBN: 0-534-09972-6] Language-independent approach to programming and problem solving, based on pseudocode, structure charts, Input/Process/Output (IPO) diagrams. New edition includes correctness issues, appendices on Fortran, BASIC, Pascal. Stresses modularity, but subprograms come very late. RM

**Programming, T(13: 1).** *Microsoft Basic: Programming the IBM PC, Second Edition.* Robert J. Bent, George C. Sethares. Ser. in Comput. Sci. Brooks/Cole, 1989, xiii + 450 pp, $32.50 (P). [ISBN: 0-534-10116-X] For the first course. A variant of the authors' other Basic texts. Semi-tutorial. Emphasis on problem solving. Many examples, problems, and programming exercises that concern diverse application areas. Includes material on menu-driven programs, sorting, data files, random numbers, graphics. Some programming exercises suitable for projects. DFA

**Programming, S(14), P, L.** *The C and UNIX Dictionary: From Absolute Pathname to Zombie.* Kaare Christian. Wiley, 1988, xii + 216 pp, $29.95. [ISBN: 0-471-60929-3] Gives all definitions of over 1000 C, UNIX system, computer and basic computer hardware terms, each listing complete with pronunciation guides, etymologies, cross-references, illustrations, derivations, and even example dialogue or code fragment where helpful. Also contains profiles of individuals who have contributed significantly to UNIX system development. DFA

**Programming, S(14), L.** *Writing Readable Ada: A Case Study Approach.* Susan Fife Dorchak, Patricia Brisotti Rice. DC Heath, 1989, xii + 244 pp, (P). [ISBN: 0-669-12616-0] To accompany any Ada text, or for self-study by someone with some familiarity with the language. Illustrates good software engineering practices and the features of Ada by developing two case studies; includes the complete code for one of them. Provides an Ada style guide and bibliography. DFA

**Programming, T*(13-14: 1).** *Programming and Problem Solving in Modula-2.* Sanford Leestma, Larry Nyhoff. Macmillan, 1989, xi + 744 pp. [ISBN: 0-02-369691-5] Designed for ACM Curriculum '84 CS 1 course using modern approach and Modula-2, plus final chapter on data structures and algorithms (stacks, queues, sorting). Standard treatment but nicely done, with good explanations, examples, exercises. RM

**Languages, P, L.** *The KornShell Command and Programming Language.* Morris I. Bolsky, David G. Korn. Prentice-Hall, 1989, xvi + 356 pp. [ISBN: 0-13-516972-0] This book is both the specification of the KornShell language and a reference handbook for **ksh**, the program that implements the KornShell language. The version of **ksh** described is consistent with Draft 6 of the POSIX standard. **ksh** is distributed as part of UNIX System V Release 4. AO

**Algorithms, T*(15-16: 1), L.** *Introduction to Algorithms: A Creative Approach.* Udi Manber. Addison-Wesley, 1989, xiv + 478 pp, $37.75. [ISBN: 0-201-12037-2] Emphasizes the creative side of algorithm design and the principles that underlie the design of new algorithms. Uses an analogy with proving mathematical theorems by induction to provide an intuitive framework for explaining a methodology of algorithm design. AO

**Algorithms, P.** *The Theory of Algorithms.* A.A. Markov, N.M. Nagorny. Transl: M. Greendlinger. Math. & Its Applic. Kluwer Academic, 1988, xxiv + 369 pp, $149. [ISBN: 90-277-2773-2] The Russian original of this book appeared in 1985. It discusses Markov's notion of a *normal algorithm* and uses this notion to investigate questions regarding the existence of algorithms. AO

**Algorithms, P, L.** *Elements of Computer Algebra with Applications.* Alkiviadis G. Akritas. Wiley, 1989, xv + 425 pp, $46.95. [ISBN: 0-471-61163-8] Focuses on the algorithms used to implement computer algebra systems (e.g., *Maple, Mathematica, MACSYMA*). Among the topics covered are error-correcting codes, computation of polynomial greatest common divisors and polynomial remainder sequences, factorization of polynomials with integer coefficients, and isolation and approximation of the real roots of polynomial equations. AO

**Computer Systems, P.** *Special Purpose Computers.* Ed: Berni J. Alder. Computat. Tech., V. 5. Academic Pr, 1988, x + 282 pp, $49.95. [ISBN: 0-12-049260-1] Seven articles by designers of particular special purpose computers, some created for efficient performance of specific numerical algorithms, others built originally as architectural examples. Computers discussed (and authors' institutions): Hypercube (Caltech); QCD machine (Rutgers); Geometry-defining processors (MIT); Navier-Stokes computer (Princeton); LCAP array (IBM); molecular dynamics processor (Delft University, The Netherlands); Delft Ising System Processor (Delft). RB

**Computer Systems, P.** *The UNIX Command Reference Guide.* Kaare Christian. Wiley, 1988, xvii + 361 pp, $27.95 (P). [ISBN: 0-471-85580-4] Subtitled "The top 50 commands: what they are, how they work, how to use them," this book is designed

to be used with either Berkeley or System V UNIX. Commands are organized into seven sections: general utility commands, file management commands, text file commands, process management commands, UNIX shells, the ex/vi text editor, and text processing commands. AO

**Computer Systems, P.** *Computer Algebra.* Ed: David V. Chudnovsky, Richard D. Jenks. Lect. Notes in Pure & Appl. Math., V. 113. Marcel Dekker, 1989, ix + 240 pp, (P). [ISBN: 0-8247-8038-8] A collection of talks and invited papers from a conference on Computer Algebra as a Tool for Research in Mathematics and Physics, held in 1984 at NYU. A long expository paper by D. Chudnovsky and G. Chudnovsky covers uses of computer algebra in research on differential and Diophantine equations. Also included is the transcript of a panel discussion, with comments from the audience, on design, applications, and scientific potential of computer algebra. PZ

**Computer Systems, C, L.** *MINIX for the IBM PC, XT, and AT.* Andrew S. Tanenbaum. Prentice-Hall, 1988, xv + 486 pp, (P). [ISBN: 0-13-584400-2] At last, an intelligent user's manual. This volume presents the workings of the MINIX operating system together with detailed instructions for its use. It can be thought of as the other half of the author's *Operating Systems* (TR, August-September 1987) which, while overlapping this volume, presents general operating systems theory. JAS

**Computer Graphics, S(16-17), P, L.** *Image Synthesis: Elementary Algorithms.* Gérard Hégron. Transl: David Beeson. MIT Pr, 1988, ix + 216 pp, $29.95. [ISBN: 0-262-08166-0] The author presents basic algorithms for image generation by computer obtained from scattered sources, systematically reviewing advantages and disadvantages of each, and provides a library of image synthesis tools for curve generation, zone filling, and geometric processing. Then he argues that two techniques, Bresenham's methods and contour following, unify the various algorithms. RB

**Artificial Intelligence, P.** *The Logic of Induction.* Halina Mortimer, I. Craig, A.G. Cohn. Transl: Ewa Such-Klimontowicz. Ser. in Artif. Intellig. Halsted Pr, 1988, 182 pp, $39.95. [ISBN: 0-470-21234-9] Translation of 1982 Polish edition by Mortimer with additional chapter by Craig on relationship of machine learning and philosophical theories of induction. Mortimer's work analyzes theories of inductive reasoning, with emphasis on interpretations of probability. KS

**Artificial Intelligence, P.** *Intelligent Decision Systems.* Samuel Holtzman. Addison-Wesley, 1989, xv + 304 pp, $37.75. [ISBN: 0-201-11602-2] An Intelligent Decision System (IDS) is an automated decision support system combining knowledge-based reasoning methods and formal methods of decision analysis. Based on mathematical "influence diagrams" (directed acyclic graphs) for modeling decisions, IDS's surpass expert systems by supplying normative power, and the ability to represent and use uncertainty. Rachel, a prototype system for assisting in decisions for infertile couples, based on a Bernoulli model of human reproduction, is discussed. RM

**Computer Science, T(17-18), P.** *Fundamentals of Pattern Recognition.* Monique Pavel. Pure & Appl. Math., V. 124. Marcel Dekker, 1989, ix + 183 pp, $89.75. [ISBN: 0-8247-8025-6] A graduate-level mathematics text in which the author sets out to formalize intuitive notions of shape, as used in applications of pattern recognition, by means of homotopy theory and shape theory (à la Borsuk), and to formalize the notions of classification and recognition in terms of category theory. Background mathematical material included; no exercises. RB

**Computer Science, S(16-18), P, L.** *A Computational Logic Handbook.* Robert S. Boyer, J. Strother Moore. Perspect. in Comput., V. 23. Academic Pr, 1988, xvi + 408 pp, $39.95. [ISBN: 0-12-122952-1] Contains updated description of logic developed by Boyer and Moore and reference guide to a related mechanical theorem-proving system. Also included are a primer for this logic as a functional programming language, proofs in this logic, uses of the logic and theorem prover, and examples. LISP code used. Appendices; references; index. RJA

**Applications, P.** *Mathematical Modeling in Combustion and Related Topics.* Ed: Claude-Michel Brauner, Claudine Schmidt-Lainé. NATO ASI Ser. E, V. 140. Martinus Nijhoff (US Distr: Kluwer Academic), 1988, xiv + 592 pp, $99. [ISBN: 90-247-2689-1] Papers from a NATO Conference in Lyon, April 1987. Topics come from the areas of detonation theory (an explosive subject), flame theory (very hot these days), turbulent combustion (always changing), complex chemistry (hard to imagine), and numerical and experimental aspects of the above. MR

**Applications, P.** *Maximum-Entropy and Bayesian Methods in Science and Engineering, Volume 2: Applications.* Ed: Gary J. Erickson, C. Ray Smith. Fund. Theories of Physics. Kluwer Academic, 1988, ix + 436 pp, $117. [ISBN: 90-277-2793-7] Twenty-seven papers from workshops held at the University of Wyoming in 1985 and at the Seattle University in 1986 and 1987. TH

**Applications (Biological Science), P.** *A Random Model for Plant Cell Population Growth.* M.C.M. de Gunst. CWI Tract, V. 58. Mathematisch Centrum, 1989, 152 pp, Dfl. 24.20 (P). [ISBN: 90-6196-365-6] Develops a mathematical model for cell population growth. Combines the ideas of the transition probability model and the Monod-population size model. The resulting model is then validated through a series of experiments. SM

**Applications (Biological Science), T(15-16), S, L.** *An Introduction to Mathematical Physiology and Biology.* J. Mazumdar. Australian Math. Soc. Lect. Ser., V. 4. Cambridge U Pr, 1989, 208 pp, $49.50; $16.95 (P). [ISBN: 0-521-37002-7; 0-521-37901-6] Interesting book written for mathematics students interested in biological applications. Topics discussed include diffusion, population dynam-

ics, biofluid mechanics, and cardiological applications. Would serve as an excellent text for a follow-up course to differential equations or as a supplement for a first course in differential equations. The examples are discussed in great detail with sufficient introduction to the biological principles. Problems follow each of ten chapters. MR

**Applications (Communication Theory), T(16-18: 1), S, P, L.** *Boolean Functions with Engineering Applications and Computer Programs.* Winfrid G. Schneeweiss. Springer-Verlag, 1989, xii + 264 pp, $59.50. [ISBN: 0-387-18892-4] Boolean functions arise in reliability theory, switching circuit theory, digital diagnostics, and communications theory—it's amazing how far you can go on only two values. This text is aimed at an engineering and computer-science audience, and includes exercises with each chapter. BC

**Applications (Engineering), T(15-16: 1), S, L.** *Mathematics for Communications Engineering.* H.B. Wood. Math. & Its Applic. Halsted Pr, 1988, 437 pp, $64.95. [ISBN: 0-470-21245-4] "Written by an engineer for engineers." For communications and electrical engineers, and students in these disciplines. Topics are of practical interest and include complex numbers, Fourier analysis, Laplace transforms, vector analysis, probability theory. Stress on physical bases. Roughly half text, one-quarter appendices on basic notions, and one-quarter exercises with answers. Text includes computer programs in BASIC. JK

**Applications (Fluid Dynamics), P.** *Mathematics in Oil Production.* Ed: Sam Edwards, P.R. King. Inst. of Math. & Its Applic. Conf. Ser., V. 18. Clarendon Pr, 1988, x + 375 pp, $75. [ISBN: 0-19-853624-0] Fourteen papers from a July 1987 conference in Cambridge, England. Primarily simulation techniques. TH

**Applications (Fluid Dynamics), P.** *Numerical Methods for Fluid Dynamics III.* Ed: K.W. Morton, M.J. Baines. Inst. of Math. & Its Applic. Conf. Ser., V. 17. Clarendon Pr, 1988, xvii + 529 pp, $95. [ISBN: 0-19-853632-1] Papers from a conference at Oxford University, March 1988. Three main themes of the conference were numerical algorithms specific to CFD, grid generation techniques, and unsteady flows. MR

**Applications (Physics), T(18: 2), S, P.** *Categories, Bundles and Spacetime Topology, Second Revised Enlarged Edition.* C.T.J. Dodson. Math. & Its Applic. Kluwer Academic, 1988, xvii + 243 pp, $79. [ISBN: 90-277-2771-6] A survey of the global features of spacetime 4-manifolds using (mostly) the category Top (topological spaces and continuous maps). Topics covered include the existence of Lorentz structures, orientability and its relation to nowhere zero differentiable 4-forms, and an account of geometric singularities via bundle completion (*First Edition,* TR, October 1980). MR

**Applications (Physics), P.** *Continua with Microstructure.* G. Capriz. Tracts in Natural Philo., V. 35. Springer-Verlag, 1989, x + 92 pp, $49. [ISBN: 0-387-96886-5] Proposes a "new general setting for theories of bodies with microstructure" such as liquid crystals and materials with voids. Four parts: general properties, special theories, thermodynamics, and mathematical problems posed by the theory. BC

**Applications (Physics), L*.** *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies, Fourth Edition.* E.T. Whittaker. Cambridge U Pr, 1988, xvii + 456 pp, $24.95 (P). [ISBN: 0-521-35883-3] Inexpensive paperback edition of the standard work on the subject. In a fascinating Foreword, Sir William McCrea discusses Whittaker, his works, the book, its evolvement through four editions, and its influence on the works of others. A "must" complement to a modern treatment of analytical dynamics. JK

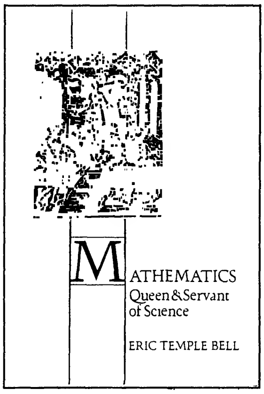
**Applications (Physics), T(15-16: 1, 2), S, L.** *Elementary Statistical Physics.* C. Kittel. Robert E Krieger, 1988, xi + 228 pp, $29.50. [ISBN: 0-89464-326-6] Reprint of a classic 1958 text. Statistical physics has advanced enormously in thirty years, but basics are basics. BC

**Applications (Social Science), P, L.** *Game Theory and National Security.* Steven J. Brams, D. Marc Kilgour. Basil Blackwell, 1988, xiii + 199 pp. [ISBN: 1-557-86003-3] A very interesting and well-written book in which game theory is applied to a wide range of problems involving the security of nations. Several deductive models are constructed. Accessible to readers with no technical background. For the interested reader, the details of most of the derivations are given in the Appendix. RH

### Reviewers

RJA: Richard J. Allen, St. Olaf; DFA: David F. Appleyard, Carleton; SB: Steve Benson, St. Olaf; RB: Richard Brown, Carleton; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; CE: Christopher Ennis, Carleton; SG: Steven Galovich, Carleton; GG: George Gilbert, St. Olaf; BH: Bruce Hanson, St. Olaf; RH: Rhonda Hatcher, St. Olaf; TH: Timothy Hesterberg, St. Olaf; JJ: Jason Jones, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JK: Joseph Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; RWN: Richard W. Nau, Carleton; GN: Gail Nelson, Carleton; AO: Arnold Ostebee, St. Olaf; SP: Samuel Patterson, Carleton; MR: Matthew Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; JS: John Schue, Macalester; SS: Seymour Schuster, Carleton; JAS: J. Arthur Seebach, Jr., St. Olaf; KS: Kay Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MS: Myriam Steinback, Macalester; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; LW: Lynne Walling, St. Olaf; PZ: Paul Zorn, St. Olaf.

# MATHEMATICS & BIOGRAPHY




## MATHEMATICS: QUEEN AND SERVANT OF SCIENCE

E.T. Bell

An absorbing account of pure and applied mathematics from the geometry of Euclid to that of Riemann and its application in Einstein's theory of relativity. The twenty chapters treat such topics as: algebra, number theory, logic, probability, infinite sets and the foundations of mathematics, rings, matrices, transformations, groups, geometry, and topology. Republished in 1987 with corrections and an added Foreward by Martin Gardner.

454 pp., ISBN 0-88385-446-3
Paperbound

List: $15.95 MAA Member: $11.95

Catalog Number QAS

*The book deserves a place in today's market. It is a much more popular work than most histories of the subject, and that is exactly what makes it accessible to students as well as to non-mathematicians. It is rewarding reading for . . . teachers and students at all mathematical levels.*

Morris Kline of The Courant Institute

## I WANT TO BE A MATHEMATICIAN
## An Automathography in Three Parts

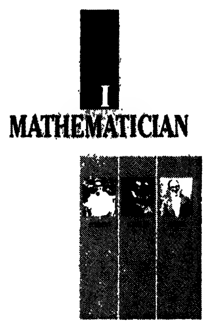
Paul R. Halmos

This is a book to be read with interest by all those who know, or might want to know, what mathematicians and mathematical careers are like. Paul R. Halmos begins with his school days and carries the reader swiftly through a career that has sustained itself at a high level since his first post-doctoral days at the Institute for Advanced Study in 1939, where he worked with John von Neumann among others. Still going strong in 1988, Halmos has contributed much to logic, operator theory, ergodic theory, and the literature in general.

442 pp., 1988, Paper, ISBN 0-88385-445-7

List: $18.00 MAA Member: $15.00

Catalog Number IWM




*It is a truly unique book, which nobody but Paul Halmos could have written. I think it will be a classic.*

Constance Reid

*The book is exciting, witty, and well worth the time invested in its study. It communicates what it means to be a mathematician.*

John Dossey in *The Mathematics Teacher*

*Order from:*

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, D.C. 20036
(202) 387-5200

# THREE IMPORTANT BOOKS ON MATHEMATICS EDUCATION

### RESPONSES TO THE CHALLENGE: KEYS TO IMPROVED INSTRUCTION BY TEACHING ASSISTANTS AND PART-TIME INSTRUCTORS

The Committee on Teaching Assistants
and Part-Time Instructors,
Bettye Anne Case, Chair

The committee that prepared this volume has been gathering information on policies, practices, successes, failures, and goals connected with the use of teaching assistants and part-time instructors. In this volume the committee presents and analyzes data showing who these teachers are, the extent and nature of their teaching duties, and the efforts made to assimilate them into the faculties. This volume will help you to see how your department compares nationally, to decide what steps you and your school should take, and to understand what additional resources might be needed.

280 pp., 1988, ISBN-0-88385-061-3
List: $15.00
Catalog Number NTE-11

### GUIDELINES FOR THE CONTINUING MATHEMATICAL EDUCATION OF TEACHERS

Committee on the Mathematical Education of Teachers

These guidelines will be very useful to school teachers and supervisors, to college administrators who plan continuing edcucation programs, to the college teachers who design and teach courses for teachers, and to school administrators who must think about requirements for continuing education of teachers. The guidelines are rich in specifics on course content, giving clear objectives for all courses. Teachers who want to dig out material for themselves or in order to enrich their classes will find the more than 500 references provided here under various topics an invaluable aid.

90 pp., 1988, ISBN-0-88385-060-5
List: $8.00
Catalog Number NTE-10

### THE USE OF CALCULATORS IN THE STANDARDIZED TESTING OF MATHEMATICS

John W. Kenelly, Editor

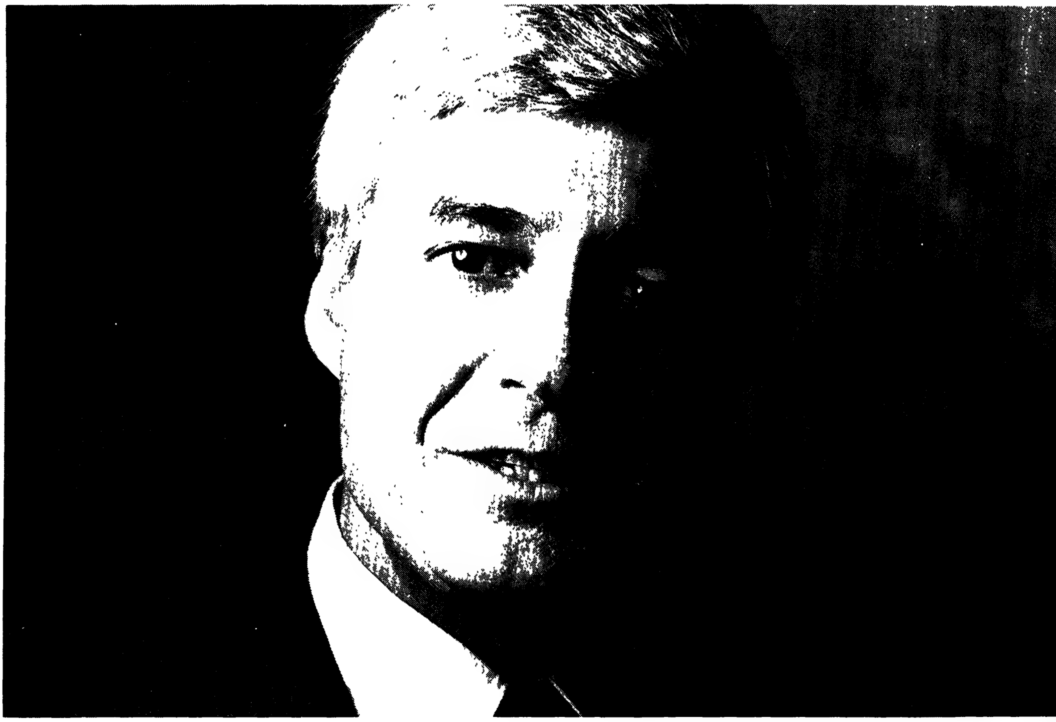
The calculator is a universal tool for all those involved in quantitative work from science and engineering to business. Routine use of calculators is part of the training and testing of students headed for these fields. But this is not yet the case in mathematics. This symposium, jointly sponsored by the MAA and The College Board, sets out clearly the theoretical and practical issues that must be addressed as calculators are brought more fully into the mathematics curriculum. This is a practical group concerned with specific tests and test items. General theoretical considerations are set off by the specifics of individual test items and students' success rates on them. The Ohio Early College Mathematics Placement Test is reported on in detail by Joan R. Leitzel and Bert K. Waits. James W. Wilson and Jeremy Kilpatrick examine the theoretical issues in the development of calculator-based tests. John Harvey, now Chair of the MAA's Committee on Placement Examinations looks at the issues surrounding calculator use on placement examinations, as well as giving an overview of the symposium and a survey of developments through 1988.

vi + 50 pp., 1989, Copublished by the MAA and The College Board. LC No. 88-064-100

List: $6.50 Member: $6.50

*Order from:*

The Mathematical Association of America
1529 Eighteenth Street, N. W.
Washington, D. C. 20036
(202) 387-5200



THE AMERICAN
# MATHEMATICAL MONTHLY

Volume 96, Number 9 November 1989

# Contents

(ISSN 0002-9890)

ARTICLES

A Survey of Transcendentally Transcendental Functions . . . . . . . . LEE A. RUBEL 777

Covering Curves by Translates of a Convex Set . . K. BEZDEK AND R. CONNELLY 789

Reconstructing a Function from Its Set of Tangent Lines . . . . . . ALAN HORWITZ 807

NOTES

Can a Graph Be Both Continuous and Discontinuous? . . . . . . HUGH THURSTON 814

Normalized Symmetric Functions, Newton's Inequalities, and a New Set of Stronger Inequalities . . . . . . . . . . . . . . . . . SHMUEL ROSSET 815

Reflection Sequences . . . . . . . . . . . . . . . N. ALON, I. KRASIKOV, AND Y. PERES 820

Counting the Rationals . . . . . . . . . . . . . . . . . . . . . . . . . . YORAM SAGHER 823

THE TEACHING OF MATHEMATICS

The Radius of Convergence of Power Series Solutions to Linear Differential Equations . . . . . . . . . . . . . . . . . . . . . . . ISOM H. HERRON 824

Pursuing Analogies Between Differential Equations and Difference Equations . . . . . . . . . . . . . . . . . . . . . DAVID L. ABRAHAMSON 827

On the Use of Iteration Methods for Approximating the Natural Logarithm . . . . . . . . . . . . . . . . . . . . . . . . . . JAMES F. EPPERSON 831

There Are No Safe Virus Tests . . . . . . . . . . . . . . . . . . . . WILLIAM F. DOWLING 835

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 837

Advanced Problems and Solutions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 846

REVIEWS

Mathematica—A System for Doing Mathematics by Computer by Wolfram Research . . . . . . . . . . . . . . . . . . . . . . . . . LAWRENCE S. KROLL 855

Dynamical Systems edited by V. I. Arnold . . . . . . . . . . . . . . JOHN HORNSTEIN 861

TELEGRAPHIC REVIEWS . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 865

# NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to ANITA E. SOLOW, Department of Mathematics, Grinnell College, Grinnell, IA 50112; in geometry or topology to DENNIS DETURCK, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Three copies of proposed problems or solutions should be sent to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, University of Calgary, Alberta, Canada T2N 1N4.




Reprint Permission: MARCIA P. SWARD, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Advertising Correspondence: Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $26 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $29. Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $90 per year.





Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Assocation of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

# A Survey of Transcendentally Transcendental Functions

LEE A. RUBEL, *University of Illinois*

LEE A. RUBEL is Professor of Mathematics at the University of Illinois at Urbana-Champaign, where he has been for thirty years, except for scholarly leaves. He got his Ph.D. at the University of Wisconsin (Madison) under R. C. Buck's direction in 1954, then spent two years at Cornell University and two years at the Institute for Advanced Study before coming to Illinois. His earlier work was mostly in Complex Variables, but his recent and current work is mostly in Algebraic Differential Equations. He has published about 160 research articles and two books, and has lectured at many universities around the world. He has been an editor of several journals, and is currently an Associate Editor of the *Monthly*.

Everybody knows that a function $f$ is called *algebraic* if it satisfies a polynomial equation $P(x, f(x)) = 0$, and is, otherwise, called transcendental.

A function (or a power series) $f$ is called *transcendentally transcendental* (TT) if it satisfies no algebraic differential equation, that is, no differential equation of the form $P(x, y, y', \dots, y^{(n)}) = 0$ where $P$ is a nontrivial polynomial in its $n+2$ variables. (We call $P$ a *differential polynomial* in this context.) If $f$ does satisfy such an equation, it is called *differentially algebraic* (DA). Differentially algebraic functions abound in elementary mathematical analysis—witness polynomials, rational functions, algebraic functions, $e^x$, $\log x$, $\sin x$, $\arctan x$, $\sec x$, Bessel functions, and so on. Moreover (see **[OST-I]**), sums, products, differences, quotients, compositional inverses, and compositions of DA functions are again DA. So a function like

$$J_0\left(\sec^{-1}\left(\frac{e^{x^2}+\sqrt{\log x}}{e^{x^2}-\sqrt{\log x}}\right)^x\right)$$

is DA, where $J_0$ is Bessel's function. Thus, one begins to wonder if there even *exist* TT functions, and what they look like.

In this survey, we exhibit many of the known TT functions, and give some indications of *why* they are TT. Knowing that a function is DA (and hence *not* TT) is knowing that it is an output of a general-purpose analog computer—this is the content of the Shannon-Pour-El-Lipshitz-Rubel Theorem (see **[SHA]**, **[POE]**, **[LIR-II]**), so that a function being TT has more than incidental significance. Our list of TT functions is surely not complete, and we apologize in advance to those people whose favorite TT function has been left out.

Historically, the first function to be proved TT was Euler's gamma function

$$\Gamma(x) = \int_0^\infty t^x e^{-t}\,dt/t.$$

This was done in 1887 by Hölder (see **[HOL]**). We will give a proof of this later, but will begin with some more conceptually immediate examples. My interpretation of

the phrase "transcendentally transcendental" (which has a marvelous ring to it!) is the following.

At first, functions were put into just two classes, "algebraic" (i.e., solutions $f(x)$ of $p(x, f(x)) = 0$, where $p$ is a nontrivial polynomial in two variables) and "transcendental," for the remaining class. Then it was realized that there are degrees of transcendentality, and the class of transcendental functions was partitioned into the "algebraically transcendental" (what we here call "differentially algebraic," but transcendental) functions, which are, from the point of view of differential algebra, only half-bad, and the "transcendentally transcendental" functions, which are just terrible. Some alternative terminology is "hypertranscendental" for TT functions, and "hypotranscendental" for DA functions. There has recently begun a line of research (see, for example, **[BAK]**, **[BAN]**, **[LAI]**, **[MOK]**, and **[STR]**) that studies even finer shades of transcendentality of $f(x)$ by asking whether it satisfies an ADE, not over the ring of polynomials in $x$, but over certain rings of entire functions of slow growth. (For example, it is shown in **[BAK]** that $\Gamma(x)$ satisfies no ADE whose (entire) coefficients $C(z)$ satisfy $|C(z)| \leqslant A\exp(o(|z|))$. This is a revealing line of research and much remains to be done, but we won't discuss it further here. Another kind of question is whether $f(x)$ satisfies an "analytic" differential equation, or whether it satisfies an algebraic difference-differential equation. We will not pursue these notions here.

We note first that to say that an analytic function $f$ (or a formal power series $f$) is DA is to say that there exists a field $F$ of finite transcendence degree over the rational numbers $\mathbb{Q}$ that contains $f$ and all its derivatives, as follows from the following result of Ritt and Gourin.

THEOREM 1. (**[RIG]**) *Let $y(x)$ be any differentially algebraic function. Then $y(x)$ satisfies an algebraic differential equation with integer coefficients.*

*Proof.* Representing by $y_p$ the $p$th derivative of $y$, we write the ADE for $y$ in the form

$$\sum c_{(i)} x^i y^{i_0} y_1^{i_1} \cdots y_n^{i_n} = 0. \tag{1}$$

Here, each $c$ is a nonzero constant. It is understood that the expressions $x^i y^{i_0} \cdots y_n^{i_n}$ are distinct from one another.

Now (1) states that, for the given function $y(x)$, the expressions $x^i y^{i_0} \cdots y_n^{i_n}$ are linearly dependent over $\mathbb{C}$, the field of complex numbers. Thus, if we set the Wronskian of these expressions equal to zero, we shall have an ADE for $y(x)$, usually of order greater than $n$, with *integer* coefficients. All that is necessary to show is that the Wronskian does not vanish identically in $x, y, \ldots, y_n$. But if it did, every function with $n$ derivatives would satisfy an equation like (1), that is, an equation with the same expressions $x^i y^{i_0} \cdots y_n^{i_n}$ that appear in (1), with constants $c$ not all zero, because for *analytic* functions (see **[KAP]**), the vanishing of the Wronskian is necessary and sufficient for linear dependence over the constants. But because we can construct a polynomial with any given values for itself and for its first $n$ derivatives at any number of points, it would follow easily that an equation like (1) exists, with coefficients $c$ not all zero, which is an identity for $x, y, y_1, \ldots, y_n$

arbitrary real numbers, which is absurd. The theorem is thus proved by contradiction.

Now let $\alpha_0, \alpha_1, \alpha_2, \ldots$ be any sequence of complex (or real) numbers that has infinite transcendence degree over $\mathbb{Q}$, and let $f(x) = \Sigma_{n=0}^{\infty}(\alpha_n/n!)x^n$. There is no difficulty in dividing the $\alpha_j$ by large integers to make the $\alpha_j$ so small (say $|\alpha_j| \leqslant 1$ for all $j$) that $f(z)$ is actually an *entire* function. Since $f^{(j)}(0) = \alpha_j$, we see that $f$ is TT. Ritt and Gourin in **[RIG]** used a similar procedure, using a diagonal argument, to produce their example.

Maillet **[MAI]** and later Mahler **[MAH-II]** used recursion relations (see below) for the coefficients of a DA power series $f(z) = \Sigma_{n=0}^{\infty}a_n z^n$ to prove that

$$|a_n| \leqslant K[n!]^{\alpha} \quad \text{for some } \alpha < \infty, \qquad K < \infty. \tag{2}$$

Therefore, to produce a TT formal power series (but convergent only at 0), we need only write, say

$$f(z) = \sum [n!]^n z^n.$$

Note that the bound (2) is sharp, since $\Sigma_{n=0}^{\infty} n! z^n$ is easily shown to satisfy the ADE

$$z^2 f'(z) + (z-1)f(z) + 1 = 0,$$

with a similar differential equation for $\Sigma[n!]^k z^n$, for any positive integer $k$.

Popken, in his 1935 thesis **[POP]**, showed that if $f(z) = \Sigma a_n z^n$ is DA, and if the $a_n$ are *algebraic* numbers, then, whenever $a_n \neq 0$,

$$|a_n| \geqslant \exp\left(-cn(\log n)^2\right),$$

where $c$ is some constant that depends on $f$. Thus,

$$\sum_{n=0}^{\infty} \frac{x^n}{(n^n)!}$$

is TT, a result that was already known to Hurwitz (see **[HUR-I]**).

A different way to produce a TT function is to produce a power series with large gaps. This was done by Ostrowski (see **[OST-I]**), Maillet (see **[MAI]**), and Popken (see **[POP]**), and later by Lipshitz and Rubel (see **[LIR-I]**). Once this is done, we can take any DA power series

$$a_0 + a_1 z + a_2 z^2 + \cdots \tag{3}$$

and introduce changes of signs to get

$$\pm a_0 \pm a_1 z \pm a_2 z^2 \pm \cdots \tag{3$'$}$$

so that the series (3′) will be TT, by making the sum of the series (3) and (3′) have large gaps. We will end this article with a better result in this direction. The gap theorem in **[OST-I]** is the following:

THEOREM 2. **[OST-I]**. *If $\Sigma_{n=0}^{\infty} f_n z^n$ is differentially algebraic and if $f_n = 0$ for $m_k \leqslant n < n_k$, for $k = 1, 2, 3, \ldots$ and if $\lim_{k\to\infty} n_k/m_k = \infty$, then $\Sigma_{n=0}^{\infty} f_n z^n$ is a polynomial.*

*Proof.* Let $f(z) = \Sigma_{k=0}^{\infty} f_k z^k$ be DA. Then the transcendence degree of the field $\mathbb{C}(z, f, f', f'', \ldots)$ over $\mathbb{C}(z)$ is finite. Hence the transcendence degree of $\mathbb{C}(f, f', f'', \ldots)$ over $\mathbb{C}$ is also finite and thus $f$ satisfies an ADE with coefficients from $\mathbb{C}$ (instead of $\mathbb{C}[z]$). This reduction will simplify our notation. Among all the ADE's with coefficients from $\mathbb{C}$, satisfied by $f$, let $F(\omega) = F(\omega, \omega', \ldots, \omega^{(m)}) = 0$ be of lowest possible order, $m$, and lowest possible total degree, $n$. $S(\omega) = \partial F/\partial \omega^{(m)}$ is the separant of $F$. By our choice of $F$, $S(f) \neq 0$, since $S$ has either lower order or lower degree in $\omega^{(m)}$ than $F$. Let

$$F(\omega) = \sum_{\kappa} a_{(\kappa)} \omega^{(\kappa_1)} \cdots \omega^{(\kappa_N)}$$

where each $a_{(\kappa)} \in \mathbb{C}$ and the sum is over those systems of integers $(\kappa) = (\kappa_1, \ldots, \kappa_N)$ with $0 \leqslant \kappa_1 \leqslant \kappa_2 \leqslant \cdots \leqslant \kappa_N \leqslant m$, $N \leqslant n$, for which $a_{(\kappa)} \neq 0$. (Recall that $m$ is the order of $F$ and $n$ the total degree of $F$.) Then, as Mahler showed in **[MAH-II]**, pp. 186–194, the $f_k$ then satisfy a recursion formula of the form

$$\alpha(k) f_k = -B(k) \sum_{(\kappa)} \sum_{[\lambda]}{}^{*} a_{(\kappa)} \frac{(\kappa_1 + \lambda_1)!}{\lambda_1!} \cdots \frac{(\kappa_N + \lambda_N)!}{\lambda_N!} f_{\kappa_1 + \lambda_1} \cdots f_{\kappa_N + \lambda_N}, \tag{4}$$

where

(i) $\alpha(k)$ is a fixed nonzero polynomial in $k$, depending on $F$ and $f$.

(ii) The first sum is over all the tuples $(\kappa)$ described above.

(iii) The second sum $\Sigma_{|\lambda|}^{*}$ is over all $N$-tuples of integers $[\lambda] = [\lambda_1, \ldots, \lambda_N]$ with $0 \leqslant \lambda_i$, for $i = 1, \ldots, N$ and with $\Sigma_{i=1}^{N} \lambda_i = h = k - m + s$. Here $m$ is the order of $F$ and $s \geqslant 0$ is a fixed integer depending on $f$ and $F$. The $*$ indicates that all the terms $f_l$ for $l \geqslant k$ are to be omitted.

(iv) $B(k) = 1$ if $k \geqslant h$, while $B(k) = h!/k!$ if $h > k$.

Suppose now that $f_l = 0$ for all $l$ with $m_j \leqslant l < n_j$, where $n_j/m_j \to \infty$. This means that we have extremely large blocks of zero coefficients, but says nothing about the nonzero coefficients. We will prove our result by induction by showing that for $j$ large, $f_{n_j} = 0$ also.

Suppose that $\alpha(k) \neq 0$ as soon as $k \geqslant k_0$. Note that

$$k + A \geqslant \sum_{i=1}^{N} (\lambda_i + \kappa_i) \geqslant k - A$$

for a fixed finite number $A$. Suppose now that $k \geqslant k_0$ and that $f_l = 0$ for $l = m_j, m_j + 1, \ldots, n_j - 1$, but $f_{n_j} \neq 0$. We write $k = n_j$. Further, suppose $m_j/n_j < \varepsilon$. But $f_{\kappa_i + \lambda_i} = 0$ as soon as $k > \kappa_i + \lambda_i \geqslant \varepsilon k$. The only way then we can have all $\kappa_i + \lambda_i$ in $\Sigma^*$ satisfy $\kappa_i + \lambda_i < \varepsilon k$ is to have $N\varepsilon k > k - A$, which is impossible if $\varepsilon$ is small enough. Thus, the theorem is proved.

In **[LIR-I]**, it was proved that if $f(z) = \Sigma_{k=1}^{\infty} f_{n_k} z^{n_k}$, where $n_k$ approaches $\infty$ faster than $\exp((\log k)^2)$, say $n_k = \exp((\log k)^{2+\varepsilon})$ for some $\varepsilon > 0$, then $f$ must be a polynomial. This includes the case of Hadamard gaps $n_{k+1}/n_k \geqslant \theta > 1$ for $k = 1, 2, 3, \ldots$. So, for example, $\Sigma_{k=0}^{\infty} a_k z^{2^k}$ is DA only if it is actually a polynomial. The idea of the proof was as follows. If $f(z) = \Sigma_{l=0}^{\infty} f_l z^l$ satisfies the gap condition (that $f_l = 0$ if $l \neq n_k$ for any $k$), so that most of the $f_l = 0$, then, for a large set of

values of $l$, only one of the terms (in the sum in the Mahler recursion formula) $f_{\kappa_1+\lambda_1} \cdots f_{\kappa_N+\lambda_N}$ (up to a permutation) is nonzero. For these values of $l$, (4) then becomes

$$\alpha(l) f_l = P(\lambda_1, \ldots, \lambda_N) f_{\kappa_1+\lambda_1} \cdots f_{\kappa_N+\lambda_N},$$

where $P$ is a fixed polynomial. Again using the fact that most of the $f_l$ are zero, one sees that $P(t_1, \ldots, t_N)$ has so many zeroes that $P \equiv 0$. The details are complicated.

Another reason that a function may be TT is that it is *universal*. For example, an entire function $f$ may have the property (see **[BIR]**, **[LUR]**, **[SEW]**, **[BLR]**, etc.) that its set of translates $f_t(z) = f(z - t)$, as $t$ runs over $\mathbb{C}$ (or even over $\mathbb{N} = (0, 1, 2, \ldots)$) is dense in the space $E$ of all entire functions, where the topology is that of uniform convergence on compact sets. This means that for any compact set $K$, any $\varepsilon > 0$, and any entire function $g$, there exists a $t \in \mathbb{C}$ such that $|f_t(z) - g(z)| < \varepsilon$ for all $z \in K$.

Let us see why such a universal $f$ must be TT. For, as remarked in the proof of Theorem 2, if $f$ satisfies an ADE, then it must satisfy an *autonomous* ADE, that is, one of the form

$$P\big(f(z), f'(z), \ldots, f^{(n)}(z)\big) = 0. \tag{5}$$

where $P$ does not explicitly involve the independent variable $z$. But then every translate $f_t$ of $f$ also satisfies (5). From elementary complex-variables theory, if a sequence $(F_n)$ of analytic functions converges uniformly on compact sets to a limit function $g$, then $F_n'$ converges likewise to $g'$, $F_n''$ to $g''$, etc. Therefore, $g$ would have to satisfy (5) if it were a limit (in this topology) of translates of $f$. But it is easy, given any one differential polynomial, to find an entire function (even a polynomial), that does not annul it.

Heins, in **[HEI]**, has constructed a universal Blaschke product in the unit disk, that is, a convergent Blaschke product

$$B(z) = \prod_{n=1}^{\infty} -\frac{\bar{z}_n}{|z_n|} \frac{z - z_n}{1 - \bar{z}_n z}$$

whose non-Euclidean translates are dense in a suitable sense. Such a Blaschke product must then be TT, by an argument similar to the one just given.

In **[VOR]** and **[REI]** it is shown that the Riemann zeta function $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$ is universal in the following sense.

THEOREM 3. **[VOR]** *Let $0 < r < 1/4$; let $f(s)$ be a function that is analytic inside the disc $\{|s| \leqslant r\}$ and continuous up to the boundary of the disc. If $f(s)$ has no zeroes inside the disc $\{|s| \leqslant r\}$, then for every $\varepsilon > 0$ there exists a real number $T = T(\varepsilon)$ such that* $\max_{|s| \leqslant r} |f(s) - \zeta(s + (\tfrac{3}{4} + iT))| < \varepsilon$.

This gives a contemporary proof that $\zeta(s)$ is TT. The first proof of this was given by Hilbert, and was written up by Stadigh (see **[STA]**) as part of his 1902 Helsinki thesis. A later proof was given by Ostrowski (see **[OST-I]**), where it was moreover shown that $\zeta(s)$ satisfies no algebraic difference-differential equation.

As we mentioned earlier, Euler's gamma function $\Gamma(x)$ was the first function to be proved TT. Hölder's proof in 1887 was based on the functional equation $\Gamma(z+1) = z\Gamma(z)$. We will give a brief proof by Ostrowski that $\Gamma(x)$ is TT. Later, Bank showed in **[BAN-II]** that $\Gamma(z)\Gamma(az)\Gamma(bz)$ is TT provided $ab \neq 0$. Other specific functions that have been shown to be TT are $\Sigma_{n=0}^{\infty} z^{2^n}$ (see **[MAH-II]**) and $\Sigma_{n=1}^{\infty}(z^n/1 - 2^n)$ (see **[CAR]**). I recall seeing somewhere (but where?) that the Bessel function $J_\nu(z)$ is TT (as a function of $\nu$).

THEOREM 4. *$\Gamma(x)$ is transcendentally transcendental.*

*Ostrowski's Proof.* (**[OST-II]**). Denote the unknown function by $y$ and its derivatives $y', y'', \ldots, y^{(\nu)}, \ldots$ by $y_1, y_2, \ldots, \underline{y}_\nu, \ldots$ . For two different power-products in $y, y_1, y_2, \ldots,$ say $A(x)y^{n_0}y_1^{n_1}y_2^{n_2} \cdots$, $\bar{A}(x)y^{\bar{n}_0}y_1^{\bar{n}_1}y_2^{\bar{n}_2} \cdots$, call the first *higher* than the second if the *last* of the nonvanishing ones among the differences $n_0 - \bar{n}_0$, $n_1 - \bar{n}_1, \ldots$ is strictly positive. This clearly defines a *transitive* ordering among the terms of a differential polynomial, so that one can talk about the *highest* term in a differential polynomial.

Among all the ADE's that $\Gamma(x)$ satisfies, supposing that it is DA, select that one whose highest term is lowest. (It is therefore of the lowest order, and of lowest degree, among those of that order (in the highest occurring derivative) and so on.) Say the equation is

$$f(y, y_1, y_2, \ldots; x) = 0$$

and the highest term is

$$A(x)y^{n_0}y_1^{n_1}y_2^{n_2} \cdots .$$

We may further assume that the degree of $A(x)$ is as small as possible and that $A(x)$ has leading coefficient 1. Then, surely, $f$ is divisible neither by $y$ nor by a linear factor $x - \alpha$.

If another ADE $\bar{f} = 0$ that is satisfied by $\Gamma(x)$ has the highest term $\bar{A}(x)y^{n_0}y_1^{n_1}y_2^{n_2} \cdots$, then $A(x)$ must be a divisor of $\bar{A}(x)$ and we have

$$\bar{f} = \frac{\bar{A}(x)}{A(x)} f.$$

For, by the Euclidean algorithm, we may write $\bar{A}(x) = QA(x) + P$, where $P$ and $Q$ are polynomials, and the degree of $P$ is lower than that of $A(x)$. Then the differential polynomial $\bar{f} - Qf$ will have a lower highest term than $f$, or else the degree of the highest term of $\bar{f} - Qf$ with respect to $x$ is smaller than the degree of $A(x)$, so that $\bar{f} - Qf$ must vanish identically.

On account of the functional equation $\Gamma(x+1) = x\Gamma(x)$, $\Gamma(x)$ also satisfies the ADE $f((x\Gamma(x)), (x\Gamma(x))', (x\Gamma(x))'', \ldots; x+1) = 0$. The left side comes from $f(y, y_1, y_2, \ldots; x)$ if one substitutes for $y, y_1, y_2, \ldots; x$ the quantities

$$\bar{y} = xy,\ \bar{y}_1 = xy_1 + y,\ \bar{y}_2 = xy_2 + 2y_1,\ \bar{y}_3 = xy_3 + 3y_1, \ldots;\ \bar{x} = x + 1. \tag{6}$$

Denote this substitution by $S$, and the result of its application to $g(y, y_1, y_2, \ldots; x)$ by $Sg$. If $B(x)y^{m_0}y_1^{m_1} \ldots$ is any power-product in $y, y_1, y_2, \ldots,$ then it is clear that

the highest term of $SB(x)y^{m_0}y_1^{m_1}\cdots$ is $x^m B(x+1)y^{m_0}y_1^{m_1}\cdots$, where $m = m_0 + m_1 + \cdots$. Hence, the highest term of $Sf$ is $x^n A(x+1)y^{n_0}y_1^{n_1}\cdots$, where $n = n_0 + n_1 + \cdots$. By the above observation, $x^n A(x+1)/A(x)$ must be a polynomial $D(x) = x^n + \cdots$, and we thus have

$$Sf \equiv D(x)f. \tag{7}$$

To invert the process to get $f$ from $Sf$, one must replace $x, y, y_1, \ldots, y_\nu, \ldots$ by

$$x-1, \frac{y}{x-1}, \frac{(x-1)y_1 - y}{(x-1)^2}, \ldots, \frac{g_\nu(y, y_1, \ldots; x)}{(x-1)^{n_\nu}}, \ldots,$$

where the $g_\nu$ are polynomials in $y, y_1, \ldots, x$, and the $n_\nu$ are positive integers. If one makes this substitution in (7) and multiplies both sides by a suitable power of $x-1$, then for a certain positive integer $a$, one obtains

$$(x-1)^a f(y, y_1, \ldots; x) \equiv D(x-1)g(y, y_1, \ldots; x),$$

where $g$ is a polynomial in $y, y_1, \ldots, x$.

Hence $D(x-1)$ can have no root $\alpha$ other than $\alpha = 1$, since otherwise $f$ would vanish for $x = \alpha$ identically as a polynomial in $y, y_1, \ldots$, and thus would have to be divisible by $x - \alpha$. Thus $D(x) = x^n$ and

$$f(xy, xy_1 + y, xy_2 + 2y_1, \ldots; x+1) \equiv x^n f(y, y_1, y_2, \ldots; x). \tag{8}$$

We set here $y = 0$ and equate the highest terms on both sides of the resulting equation

$$f(0, xy_1, xy_2 + 2y_1, \ldots; x+1) \equiv x^n f(0, y_1, y_2, \ldots; x).$$

If the highest term of $f(0, y_1, y_2, \ldots; x)$ is, say, $C(x)y_1^{l_1}y_2^{l_2}\cdots$, then we have

$$x^{l_1+l_2+\cdots}C(x+1)y_1^{l_1}y_2^{l_2}\cdots \equiv x^n C(x)y_1^{l_1}y_2^{l_2}\cdots.$$

Consequently we must have $l_1 + l_2 + \cdots = n$ and $C(x+1) = C(x)$, so that $C(x)$ is a nonzero constant $C$. Consequently $f(0, y_1, y_2, \ldots; x)$ is not divisible by $x-1$ and $f(0, y_1, y_2, \ldots; 1)$ does not vanish identically. But if we put $x = 0$ in (8), then

$$f(0, y, 2y_1, 3y_2, \ldots; 1) \equiv 0,$$

or, replacing $y, 2y_1, 3y_2, \ldots$ by $y_1, y_2, y_3, \ldots$, respectively, we get

$$f(0, y_1, y_2, \ldots; 1) \equiv 0,$$

which contradicts what we have just proved, and thus $\Gamma(x)$ must be transcendentally transcendental.

In **[RIT-II]**, Ritt found all DA solutions of the Poincaré functional equation (see **[POI]**)

$$y(mx) = R[y(x)], \tag{9}$$

with $y$ a meromorphic function in $\mathbb{C}$ and $R(x)$ a rational and not linear function. He showed that if $y(x)$ is such a DA solution of (9), then it is obtained directly from one of the functions $e^{ax}, \cos(\alpha x + \beta)$, or the Weierstrass $p$-function $p(x)$. This is one way to show that certain functions are TT.

Perhaps surprisingly, the Jacobi theta function $\theta(x) = \Sigma_{n=0}^{\infty} x^{n^2}$ is actually DA. This fact is proved periodically—see **[JAC]**, **[HUR-II]**, **[DRA]**, **[RES]**, **[CHO]**, occurring with the dates 1847, 1889, 1938, 1966, and 1984, respectively, for a mean period of 35 years. This result shows, in conjunction with **[ALE]**, that there are DA imbeddings of the unit disc in $\mathbb{C}^1$ into $\mathbb{C}^2$, which comes as a surprise, at least to me. Nobody seems to know whether $\theta_3(x) = \Sigma_{n=0}^{\infty} x^{n^3}$ is TT. In **[BOR]** and **[CRR]**, solutions of certain Schröder's functional equations were shown to be TT.

In **[STE]**, Steinmetz proved that if $f$ and $g$ are *entire* nonconstant functions, and if the composition $f \circ g$ is DA, then both $f$ and $g$ must be DA. Thus, $f(1/\Gamma(z))$ and $1/\Gamma(g(z))$ are always TT if $f$ and $g$ are non-constant entire functions. In **[BOS]**, it was shown that any analytic (in a disc) function $\phi(z)$ such that $\phi(\phi(z)) = e^z$ must be TT.

Of course, one way to produce TT functions of a real variable is by insufficient differentiability. Weierstrass's nowhere-differentiable function is an extreme example. Less trivially, it was shown in **[BRR]** that if $f(x)$ satisfies an ADE (with the only differentiability requirement being that $f$ have enough derivatives to enable it to be plugged into the ADE), then $f$ must be analytic on a dense open subset of its domain. Using this, one can produce many TT functions.

Ostrowski, in **[OST-I]**, produced real-analytic functions $f(x, y)$ that satisfy no algebraic *partial* differential equations (APDE). Here's an intriguing problem. If $u(x, y)$ satisfies an APDE and $v(x, y)$ satisfies an APDE, must $u(x, y) + v(x, y)$ also satisfy an APDE? (Suppose, say, that $u$ and $v$ are real-analytic.) NOTE: Since this was written, Wolfgang Schmidt has sent the author a lovely proof of the affirmative answer to this problem.

We close this survey with two "new" results whose gist is that "most" entire functions are TT.

Let $E$ be the space of all entire functions, in the topology of uniform convergence on compact sets. It is known (see **[LUR]**) that $E$ is a complete metric space.

THEOREM 5. *The differentially algebraic entire functions form a set of the first Baire category in $E$.*

*Remark.* There is no trouble in extending this result (and its proof) from $E$ to $H(G)$, the space of all holomorphic functions on a region $G$ in $\mathbb{C}$.

*Proof of Theorem* 5. By Theorem 1 and an early remark, if $f \in$ DA, then $f$ must annul one of $P_1, P_2, P_3, \ldots,$ which are the *autonomous* nontrivial differential polynomials with integer coefficients, written out in a countable list. So it is enough to prove that each $P_j^{\perp} = \{f \in E : P_j \text{ annuls } f\}$ is nowhere dense. But it is clear that each $P_j^{\perp}$ is closed in $E$ (since if $f_j \to f$ in $E$, then $f_j^{(m)} \to f^{(m)}$ in $E$ for $m = 0, 1, 2, \ldots$.) Hence, it is enough to prove that each $P_j^{\perp}$ contains no neighborhood in $E$. But a basic neighborhood, $N(f: K, \varepsilon)$ of the function $f \in E$ is indexed by the compact set $K \subseteq \mathbb{C}$ and the positive number $\varepsilon$, and is given by

$$N(f: K, \varepsilon) = \{g \in E : \sup\{|f(z) - g(z)| : z \in K\} < \varepsilon\}.$$

Given $f$, $K$, $\varepsilon$, look at

$$g(z) = f(z) + \sum_{k=0}^{M} \frac{\varepsilon_k}{k!} z^k. \tag{10}$$

If all the $\varepsilon_k$ are chosen to be sufficiently small complex numbers, say $|\varepsilon_k| < \varepsilon'$, $k = 0, 1, \ldots, M$, then $g \in N(f\colon K, \varepsilon)$. Now let $P(\not\equiv 0)$ be an autonomous differential polynomial. We claim that $P^{\perp}$ contains no $N(f\colon K, \varepsilon)$. Otherwise, we would have $P(g) = 0$ for every $g$ of the form (10), with the $\varepsilon_k$ small enough. But $g^{(l)}(0) = f^{(l)}(0) + \varepsilon_l$, for $l = 0, 1, \ldots, M$, so that $P(f(0) + \varepsilon_0, f'(0) + \varepsilon_1, \ldots, f^{(n)}(0) + \varepsilon_n) = 0$. Thus $P(w_0, w_1, \ldots, w_n)$ would vanish identically over a whole open ball in $\mathbb{C}^{n+1}$, and thus $P \equiv 0$, contradictory to our hypothesis.

In an earlier version of this survey, we stated and proved the following (true) result.

THEOREM 6. *Given a power series $\sum_{n=0}^{\infty} a_n z^n$ that is not a polynomial, the probability is* 1 *that the randomized series $\sum_{n=0}^{\infty} \pm a_n z^n$ is transcendentally transcendental.*

*Discussion.* The assertion means that the set of $\pm$ for which $\sum_{n=0}^{\infty} \pm a_n z^n$ is DA has measure 0, where to a sequence of $\pm$ signs, say $((-1)^{\varepsilon_n})$, $\varepsilon_n = 0$ or 1, we assign the real number $\sum_{n=0}^{\infty} \varepsilon_n 2^{-(n+1)}$ in the interval $[0, 1]$, and the measure is ordinary Lebesgue measure. But much more than this is true—namely that there are at most *countably* many choices of $\pm$ for which $\sum_{n=0}^{\infty} \pm a_n z^n$ is differentially algebraic. This is a consequence of the next result, which is essentially equivalent to Proposition 6.1 of [**LIR-I**], that there are only countably many *spectra* for the totality of differentially algebraic power series. (The spectrum of $\sum_{k=0}^{\infty} f_k z^k$ is the set of $k$ where $f_k \neq 0$.) That result is due to Laohakosol, Lipshitz, Richman, and the author, jointly.

THEOREM 6*. *Given a (finite or) countable set $S$, there are at most countably many differentially algebraic power series all of whose coefficients are elements of $S$.*

One gets the first-mentioned result from Theorem 6* by choosing $S = \{a_0, -a_0, a_1, -a_1, a_2, -a_2, \ldots\}$. Our proof depends on the Hurwitz recursion formula, which we now state in detail. (For a proof, see [**DEL**], specifically the proofs of Lemmas 2.3 and 2.4.)

LEMMA H. (*The Hurwitz recurrence relation.*) *Let $P(x, y, y', \ldots, y^{(n)})$ be any differential polynomial with coefficients in $\mathbb{Q}$ in the differential indeterminate $y$, of order $n$. Then for any $k \in \mathbb{N}$, $P^{(2k+2)} = y^{(n+2k+2)} f_n + y^{(n+2k+1)} f_{n+1} + y^{(n+2k)} f_{n+2} + \cdots + y^{(n+k+2)} f_{n+k} + f_{n+k+1}$, where the $f_j$ are differential polynomials (with rational coefficients) in $y$ of order at most $j$, for $j = n, n+1, \ldots, n+k+1$, and $f_n = \partial P / \partial y^{(n)}$. Note that $f_{n+1}, f_{n+2}, \ldots$ depend on $k$. Let $\bar{y} \in \mathbb{C}[[x]]$, the formal power series in $x$, satisfy $P(x, \bar{y}, \bar{y}', \ldots, \bar{y}^{(n)}) = 0$, and suppose that*

$$\frac{\partial P}{\partial y^{(n)}}(x, \bar{y}, \bar{y}', \ldots, \bar{y}^{(n)}) = C_0 x^k + C_1 x^{k+1} + \cdots \tag{11}$$

*with* $C_0 \neq 0$. *Then there exists a least* $r \in \mathbb{N}$, $0 \leqslant r \leqslant k$, *such that*

$$\left[f_{n+r} + qf'_{n+r-1} + \cdots + \binom{q}{r} f_n^{(r)}\right](0, \bar{y}(0), \bar{y}'(0), \dots) \tag{12}$$

*is a nonzero polynomial in* $q$. *Let* $\gamma \in \mathbb{N}$ *be bigger than any root* $q \in \mathbb{N}$ *of polynomial* (12). *Then for all* $q \geqslant \gamma + r$,

$$\bar{y}^{(n+2k+2+q-r)}(0) = \frac{-H_{n+2k+1+q-r}(0, \bar{y}_0, \bar{y}'_0, \dots)}{A(0, \bar{y}_0, \bar{y}'_0, \dots, q)} \tag{13}$$

*where* $A(x, y, y', \dots, q) = f_{n+r} + qf'_{n+r-1} + \cdots + \binom{q}{r} f_n^{(r)}$, *and* $H_{n+2k+1+q-r}$ *is a rational-coefficient differential polynomial in* $y$ *of order at most* $n + 2k + 1 + q - r$, *whose definition depends only on* $P$, $k$, $\gamma$, $r$, *and* $q$.

We remark that the ADE $y(xy)'' - (xy)'y' = 0$ has the solution $y = ax^l$ for any positive integer $l$ and any constant $a$. This example shows that the recurrence relation for the coefficients of the solutions of an ADE may well depend on the solution as well as on the ADE.

*Proof of Theorem* 6*. If $\phi(z) = \sum_{n=0}^{\infty} s_n z^n$ is DA for uncountably many choices of $s = (s_n)$, where all the $s_n \in S$, then by Theorem 1, there must be a differential polynomial $P$ with *integer* coefficients that annuls uncountably many of the $\phi$. On taking the successive separants $S_1, S_2, S_3, \dots$ of $P$, we eventually arrive at a non-zero constant. (Here, $S_1 = \partial P/\partial y^{(n)}$, where $n$ is the order of $P$, and $S_{j+1} = \partial S_j/\partial y^{(n_j)}$, where $n_j$ is the order of $S_j$.)

Thus, we would have, for an uncountable set of $\phi$, that there is a fixed differential polynomial $P$ that annuls $\phi$, such that its separant $S_p$ does not annul $\phi$. Turning now to Lemma H, we will have an uncountable subset of $\phi$ that shares the same $k$. Then there will be an uncountable subset of that set that shares the same $r$. Then there will be an uncountable subset of that set, where the biggest integer root of the polynomial (12) is $\leqslant L$ for some fixed integer $L$. Here, $k$ and $r$ are the quantities appearing in Lemma H, and we take $\bar{y} = \phi$. On this last set of $\phi$'s we have, for $N \geqslant N_0$,

$$\frac{\phi^{(N+1)}(0)}{(N+1)!} = \frac{-H_N(0, \phi(0), \phi'(0), \dots)}{A(0, \phi(0), \phi'(0), \dots, q)}.$$

This says, that for these $\phi$'s, the coefficients are determined once we know the first $N_0 + 1$ of them. Since there are only countably many choices of the initial segment of coefficients, our supposedly uncountable set turns out to be countable after all. This contradiction proves the Theorem.

We end the paper with an

OUTRAGEOUS CONJECTURE. *Let* $\varepsilon_n = 0, 1$ *for* $n = 0, 1, 2, \dots$ . *Then* $\sum_{n=0}^{\infty} \varepsilon_n z^n$ *is a differentially algebraic function if and only if* $\sum_{n=0}^{\infty} \varepsilon_n 2^{-n}$ *is an algebraic number.*

Since the conjecture is officially labelled as "outrageous" we don't have to justify it except to say that we haven't found any obvious counterexamples.

REFERENCES


[ALE] Herbert Alexander, Explicit imbedding of the (punctured) disc into $\mathbb{C}^2$, *Comment. Math. Helvetici*, 52 (1977) 539–544.

[BAK] Steven B. Bank and Robert P. Kaufman, An extension of Hölder's theorem concerning the gamma function, *Funkcialaj Ekvacioj*, 19 (1976) 53–63.

[BAN-I] Steven B. Bank, On algebraic differential equations whose coefficients are entire functions of finite order, *Rev. Roumaine Math. Pures Appl.*, 20 (1975) 511–519.

[BAN-II] ______, On certain canonical products which cannot satisfy algebraic differential equations, *Funkcialaj Ekvacioj*, 23 (1980) 335–349.

[BIR] George David Birkoff, Demonstration d'un théorème élémentaire sur les fonctions entières, *C. R. Acad. Sci. Paris*, 189 (1929) 473–475.

[BLR] Charles Blair and Lee A. Rubel, A triply universal entire function, *l'Enseignement Math.*, 30 (1984) 269–274.

[BOR] Michael Boshernitzan and Lee A. Rubel, Coherent families of polynomials, *Analysis*, 6 (1986) 339–389.

[BOS] Michael Boshernitzan, Hardy fields, existence of transexponential functions, and the hypertranscendentality of solutions to $g(g(x)) = e^x$, preprint.

[BRR] Andrew M. Bruckner, Melvin Rosenfeld, and Lee A. Rubel, The Darboux property and solutions of algebraic differential equations, *Pacific J. Math.*, 118 (1985) 295–302.

[CAR] R. D. Carmichael, On transcendentally transcendental functions, *Trans. Amer. Math. Soc.*, 14 (1913) 311–319.

[CHO] P. Chowla and S. Chowla, On the algebraic differential equations satisfied by some elliptic functions II, *Hardy-Ramanujan Journal*, 7 (1984) 13–16.

[CRR] Francis W. Carroll, Transcendental transcendence of solutions of Schröder's equation associated with finite Blaschke products, *Michigan Math. J.*, 32 (1985) 47–57.

[DAR] Jules Drach, Sur l'équation différentielle du troisième ordre des fonctions theta elliptiques, *C. R. Acad. Sci. Paris*, 206 (1938) 1421–1424

[DEL] J. Denef and L. Lipshitz, Power series solutions of algebraic differential equations, *Math. Ann.*, 267 (1984) 213–238.

[HEI] Maurice Heins, A universal Blaschke product, *Arch. Math.* (Basel) 6 (1955) 41–44.

[HIL] David Hilbert, Mathematische Probleme, Gött. Nachr. 1900, p. 253–297.

[HOL] Otto Hölder, Über die Eigenschaft der Gamma Funktion keiner algebraische Differentialgleichung zu genügen, *Math. Ann.*, 28 (1887) 1–13.

[HUR-I] Adolf Hurwitz, Sur le développement des fonctions satisfaisant à une équation différentielle algébrique, *Ann. Ecole Norm. Sup.*, 6 (1889) 327–332.

[HUR-II] ______, Über die Differentialgleichungen dritter Ordnung, welchen die Formen mit linearen Transformationen in sich genügen, *Math. Ann.*, 33 (1889) 345–352.

[JAC] D. G. C. Jacobi, Über die Differentialgleichung welche die Reihe... genügen leisten, *J. reine angew. Math.*, 36 (1847) 97–112.

[KAP] Irving Kaplansky, An Introduction to Differential Algebra, second edition, Hermann, Paris, 1976.

[LAI] Ilpo Laine, On the behaviour of the solutions of some first-order differential equations, *Ann. Acad. Sci. Fenn.*, Ser. A1, 497 (1971) 1–26.

[LIR-I] Leonard Lipshitz and Lee A. Rubel, A gap theorem for power series solutions of algebraic differential equations, *Amer. J. Math.*, 108 (1986) 1193–1214.

[LIR-II] ______, A differentially algebraic replacement theorem, and analog computability, *Proc. Amer. Math. Soc.*, 99 (1987) 367–372.

[LUR] Daniel Luecking and Lee A. Rubel, Complex Analysis, a Functional Analysis Approach, Springer, 1984.

[MAH-I] Kurt Mahler, Arithmetische Eigenschaften einer Klasse Transzendental-transzendenten Funktionen, *Math. Z.*, 32 (1930) 545–585.

[MAH-II] ______, Lectures on Transcendental Numbers, Lecture Notes in Mathematics, vol. 546, Springer, 1976.

[MAI] Edmond Maillet, Sur les séries divérgentes et les équations différentielles, *Ann. Sci. Ecole Norm. Sup.*, Ser. 3, 3 (1903) 487–518.

[MOK] A. Z. and V. D. Mokhon'ko, Estimates of the Nevanlinna characteristics of certain classes of meromorphic functions, and their applications to differential equations, *Siberian Math. J.*, 15 (1974) 921–934.

[MOO] Eliakim Hastings Moore, Concerning transcendentally transcendental functions, *Math. Ann.*, 48 (1896–97) 49–74.

[OSG] Charles F. Osgood, Differential algebra and the simultaneous approximation by rational functions of power series solutions to algebraic functional equations, *Amer. J. Math.*, 103 (1981) 469–497.

[OST-I] Alexander Ostrowski, Über Dirichletsche Reihen und algebraische Differentialgleichungen, *Math. Z.*, 8 (1920) 241–298.

[OST-II] ______, Zum Hölderschen Satz über $\Gamma(x)$, *Math. Annalen*, 94 (1925) 248–251.

[PAS] P. I. Pastro, The $q$-analogue of Hölder's theorem for the gamma function, *Rend. Sem. Mat. Univ. Padova*, 70 (1983) 47–53.

[POE] Marian Boykan Pour-El, Abstract computability and its relation to the general-purpose analog computer, etc., *Trans. Amer. Math. Soc.*, 199 (1974) 1–28.

[POI] Henri Poincaré, Sur une classe nouvelle de transcendantes uniformes, *Journal de Mathématiques*, ser. 4, VI (1890) 313–365.

[POP] Jan Popken, Über arithmetische Eigenschaften analytischer Funktionen, North-Holland Press, Amsterdam, 1935.

[REI] Axel Reich, Zeta funktionen und Differenzen-Differentialgleichungen, *Arch. Math.* (Basel), 38 (1982) 226–235.

[RES] H. L. Resnikoff, On differential operators and automorphic forms, *Trans. Amer. Math. Soc.*, 124 (1966) 334–346.

[RIG] Joseph Fels Ritt and Eli Gourin, An assemblage-theoretic proof of the existence of transcendentally transcendental functions, *Bull. Amer. Math. Soc.*, 33 (1927) 182–184.

[RIT-I] Joseph Fels Ritt, Periodic functions with a multiplication theorem, *Trans. Amer. Math. Soc.*, 23 (1922) 16–25.

[RIT-II] ______, Transcendental transcendency of certain functions of Poincaré, *Math. Ann.*, 95 (1926) 671–682.

[RUS] Lee A. Rubel and Michael F. Singer, A differentially algebraic elimination theorem with applications to analog computability in the calculus of variations, *Proc. Amer. Math. Soc.*, 94 (1985) 653–658.

[SEW] W. Seidel and J. L. Walsh, On approximation by Euclidean and non-Euclidean translations of an analytic function, *Bull. Amer. Math. Soc.*, 47 (1941) 916–920.

[SHA] Claude Shannon, Mathematical theory of the differential analyzer, *J. Math. Phys. Mass. Inst. Tech.*, 20 (1941) 337–354.

[STA] V. E. E. Stadigh, Ein Satz ueber Funktionen die algebraische Differentialgleichungen befriedigen und ueber die Eigenschaft der Funktion $\zeta(s)$ keiner solchen Gleichung zu genügen, thesis, Helsinki, 1902.

[STE] Norbert Steinmetz, Über die faktorisierbaren Lösungen gewöhnlicher Differentialgleichungen, *Math. Z.*, 170 (1980) 169–180.

[STR] Shlomo Strelitz, Asymptotic properties of entire transcendental solutions of algebraic differential equations, Value Distribution Theory and its Applications, New York, 1983, 171–214.

[TIE] Heinrich Tietze, Über Funktionalgleichungen, deren Lösungen keiner algebraischen Differentialgleichungen genügen können, *Monatshefte Math. u. Phys.*, 16 (1905) 329–364.

[VOR] S. Voronin, A theorem on the "universality" of the Riemann zeta-function, *Math. U.S.S.R.-Izv.*, 9 (1975) 443–453.

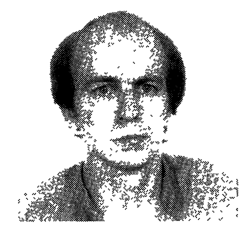
[WIT] Hans Wittich, Ganze Transzendente Lösungen algebraischen Differentialgleichungen, *Math. Ann.*, 122 (1950) 221–234.

[YAN] Niro Yanagihara, Hypertranscendancy of solutions of some difference equations, *Japan. J. Math.* (N. S.), 7 (1981) 109–168.

# Covering Curves by Translates of a Convex Set

K. Bezdek, *Eötvös University*

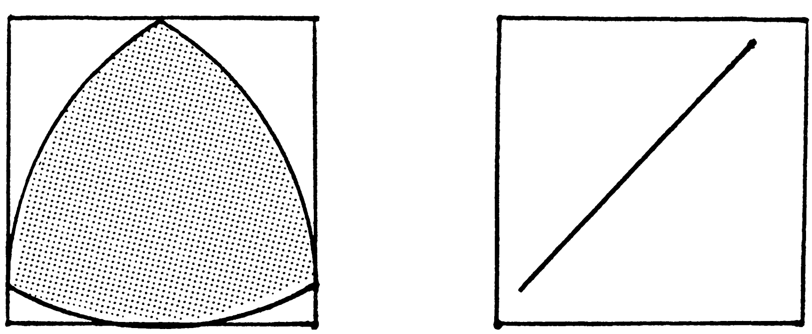
R. Connelly*, *Cornell University*



Károly Bezdek: I received my degrees in Budapest (Hungary), the M.S. Degree in 1978 and the University Doctorate Degree in 1980 at Eötvös Loránd University, the Kandidátus Degree (Ph.D.) in 1985 from the Hungarian Academy of Sciences. I have been teaching at the Department of Geometry of Eötvös Loránd University since 1978. I was on leave in the academic years 1985–86 and 1988–89 visiting Cornell University in Ithaca. My research focuses on different problems of discrete and convex geometry.

Robert Connelly: I received my Ph.D. in topology at the University of Michigan in 1969. Since then I have been at Cornell University on and off with excursions to I.H.E.S. in France, Syracuse University, the Max Planck-Institut für Math. in Bonn, West Germany, the University of Dijon and Savoie University in France, Eötvös University in Budapest, Hungary, and the Center for Research at the University of Montreal, Canada. My interests were compulsively attracted to geometry in the mid-1970's and since then I have been interested in the rigidity of frameworks and surfaces, reconstructing asteroid shapes, simplexwise-linear homeomorphisms of a two-disk, and the rigidity of packings.

**I. Introduction.** Turning a key in a lock is a familiar experience. Imagine a "key" that is a cylinder with convex planar cross-section. Imagine a "lock" which is a cylindrical hole with another convex cross-section. When can such a key fit inside the lock and turn 360°? Clearly for this problem we need only consider the cross-sections of the lock and key. For instance, the shaded set of Figure 1a can turn completely 360° inside the square, but the line segment of Figure 1b is longer than the length of a side of the square so it can never be rotated inside the square so as to be parallel to one of the sides of the square.



(a) (b)

Fig. 1

*Partially supported by NSF grant number MCS-790251.

Being mathematically curious, one might ask the following related basic question: What geometric conditions will insure that a convex set can be translated into another? For instance Wetzel [**8**] shows that for a given acute triangle, if a closed curve has length equal to or less than the length of the perimeter of the pedal triangle, then the closed curve can be translated into the given acute triangle. The pedal triangle is the triangle formed by the three feet of the altitudes of the triangle. See Figure 2.

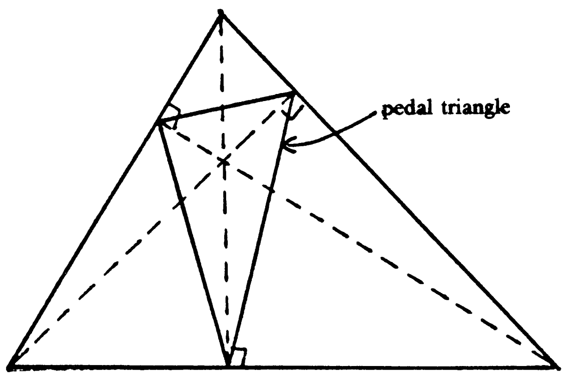



FIG. 2

Since rotating a curve does not change its length, any of the curves in Wetzel's Theorem can be rotated inside the given triangle. (It turns out for compact planar sets $C$ and $X$, $X$ convex, if for every $0 \leqslant \theta \leqslant 360°$, $C$ rotated by $\theta$ can be translated into $X$, then $C$ can be translated and then continuously rotated 360° inside $X$.)

Another example is when the covering set is a circular disk of diameter $1/2$. Then any closed curve of length one or less can be translated into the disk. This is an old result that can be found in the standard reference in this area, Bonnesen and Fenchel [**2**, p. 82]. (See also Nitche [**14**].) Note that since the covering set is a circular disk, the problem reduces simply to finding some congruent copy of the set to be covered in the disk. Any key can be turned in a round lock.

In section II we collect some principles and techniques that can possibly reduce the problem (of covering every element of a collection of sets by a translate of a fixed convex set) to a routine exercise.

Wetzel [**21**] looked at the problem of finding a plane convex set of minimum area that can cover any closed curve of length one or less with a translate. We do not know the answer to this problem either, but in section VI we find another set, with smaller area than the ones that Wetzel found, that can cover any closed curve of length one or less with a translate.

Another natural problem is to find a plane convex set of minimum perimeter that can cover any closed curve of length one or less with a translate. Our Theorem below answers this problem.

A line is called a *support line* for a compact set if the line has a non-empty intersection with the set, and the set is contained in one of the closed half planes

with the line as boundary. A compact convex set has *constant breadth* $b$ if the distance between any pair of distinct parallel support lines is the constant $b$.

We will prove the following:

THEOREM. *Let $X$ be any compact convex set of constant breadth $1/2$ in the plane, and let $C$ be a closed curve of length one or less in the plane. Then $C$ can be covered by a translate of $X$. Furthermore, if $Y$ is any compact convex set such that every closed curve of length one or less can be covered by a translate of $Y$, then the length of the perimeter of $Y$ is equal to or larger than $\pi/2$ with equality if and only if $Y$ has constant breadth $1/2$.*

*Remark* 1. An easy compactness argument shows that there must be a compact convex set $X$ of minimum perimeter such that any closed curve in the plane with length one or less can be covered by a translate of $X$. What is surprising is that any compact convex set of constant breadth will serve as such a minimal $X$. A famous theorem of Barbier [**1**] says that all compact convex sets of constant breadth, $1/2$ say, have perimeters of the same length $\pi/2$, so they all are at least candidates.

Figure 3 shows some examples of the Theorem. Each of the sets on the top is a convex set of the same constant breadth and each curve on the bottom has length less than twice that breadth. Thus each curve on the bottom can be translated into each set on the top (and then turned 360° inside it).

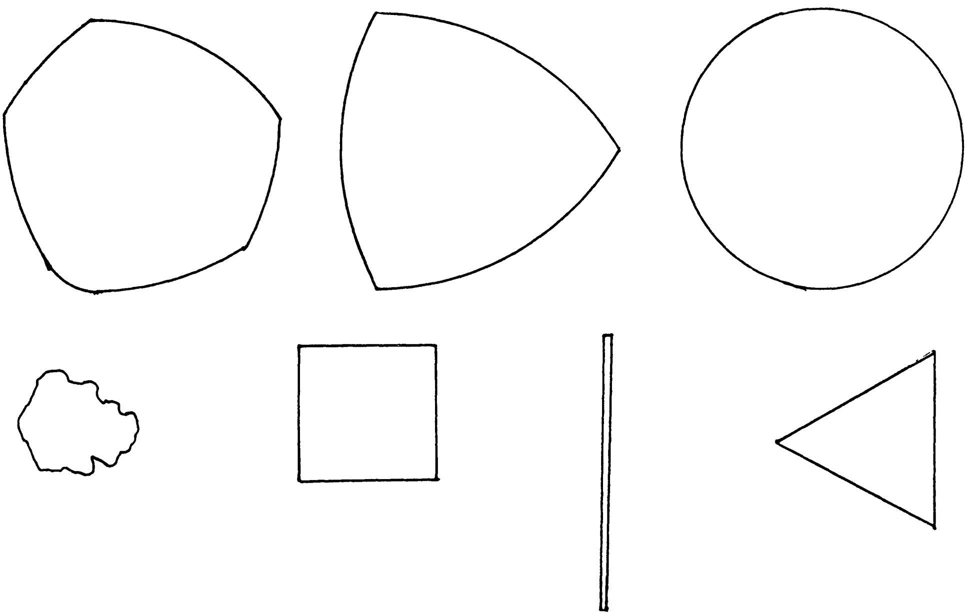


FIG. 3

Let $\mathscr{C}$ be any collection of sets in the plane. We say a set $X$ is a *translation cover* for $\mathscr{C}$ if every set in $\mathscr{C}$ can be covered by a translate of $X$. Let $\mathscr{L}_1$ be the collection of closed curves in the plane of length one or less. Our theorem then can be restated as saying that the compact convex sets of constant breadth $1/2$ are all the translation covers for $\mathscr{L}_1$ with minimum perimeter.

The terminology "translation cover" comes from Wetzel [21, p. 368]. If a compact convex set $X$ can cover any element of $\mathscr{C}$ by some congruent copy, then Wetzel calls $X$ a *displacement cover* for $\mathscr{C}$. Eggleston [6, p. 131], uses the word "universal cover" instead of displacement cover. Chakerian [3, p. 759], and Lay [12, p. 85] use the word "strong universal cover" instead of translation cover. A displacement cover is just a lock that allows the key to enter, but the key might not be able to turn.

We prefer Wetzel's terminology. However, Chakerian and Lay have an interesting discussion of such covers, whatever they are called, as well as other related subjects. The reader is also referred to [4], [5], and [17]–[21] for further information about more related covering problems.

In the following proof of the main Theorem (which occupies sections II-V of this paper) the exposition will be self-contained, except for three results: Helly's Theorem (Lemma 1, below), Sperner's Lemma (Lemma 5, below), and Fagnano's Theorem about pedal triangles (Lemma 7, below). These are all basic results that have been amply discussed elsewhere, and we have nothing to add to their proofs. However, we can apply a result of Chakerian to shorten the last step of the proof of our main Theorem. Nevertheless, we will also provide another proof, without using Chakerian's result, following our ideas of using billiard triangles.

Section II contains a series of general lemmas that, roughly speaking, reduce the covering problem to covering triangles, and the "local" problem to the case of covering a triangle by a triangle. In our case this turns out to be the billiard condition as discussed in section III. This in turn gives us a way of calculating a number for a convex set $X$ that is the upper bound for the length of a closed curve to insure that it can be covered by a translate of $X$. This is applied in section IV to the Reuleaux triangle, in section V to prove the main Theorem, and in section VI to investigate Wetzel's problem.

In section VII we look at other related problems and generalizations regarding translation covers, and mention several more results, without proof, that we can obtain with our techniques.

We thank the referees for many helpful comments, especially for informing us of Chakerian's result, and suggestions regarding an earlier version of this paper. We also thank John Wetzel for referring us to his papers [18], [20], and [21], and Maria Terrell for helpful advice.

**II. General Helly-type facts.** There are many reductions that are very helpful. We start with Helly's Theorem.

LEMMA 1 (Helly). *Let $\mathscr{C}$ be any collection of compact convex sets in the plane. If every* 3 *sets of $\mathscr{C}$ have a nonempty intersection, then the intersection of all the sets of $\mathscr{C}$ have non-empty intersection.*

A good reference for the many proofs and generalizations of this result is Danzer, Grünbaum, Klee [5].

For any set $X$ in the plane, and a point $p$ in the plane, define

$$p + X = \{ p + x | x \in X \}.$$

the *translate* of $X$ by $p$. All points in the plane are regarded as vectors in some coordinate system.

The following observation changes our covering problem to an intersection problem.

LEMMA 2. *Let $X$ and $Y$ be subsets of the plane. A translate of $X$ contains $Y$ if and only if*

$$\bigcap_{y \in Y} (-y + X) \neq \varnothing.$$

*Proof.* $p \in \bigcap_{y \in Y}(-y + X)$ if and only if $p \in (-y + X)$ for all $y \in Y$ if and only if $p + y \in X$ for all $y \in Y$ if and only if $p + Y \subset X$. □

We can now reduce the covering problem to the case of 3 points. This can be found in Chakerian [3] and is due to Klee [11]. The next Lemma follows immediately from Lemma 1 and Lemma 2.

LEMMA 3. *Let $Y$ be any set in the plane with at least three points. A compact convex set $X$ has a translate which covers $Y$ if and only if $X$ has a translate which covers every three-point subset of $Y$.*

*Proof.* By Lemma 2 a translate of $X$ covers $Y$ if and only if $\bigcap_{y \in Y}(-y + X) \neq \varnothing$. Since every three-point subset of $Y$ has a translate of $X$ which covers it, every three of the sets in the intersection have a nonempty intersection. By Lemma 1 the whole intersection is nonempty. Thus Lemma 3 follows. □

*Remark* 1. Since any three points on a closed curve of length one or less form a triangle of perimeter one or less, by Lemma 3 we can assume, without loss of generality, that the curve we want to cover is a triangle with the understanding that we allow the triangle to possibly degenerate to a line segment.

To ease the development we introduce a bit of notation. A *triangle* in the plane is written as $\Delta = \langle p_1, p_2, p_3 \rangle$, where $p_1, p_2, p_3$ are the vertices of $\Delta$. Note that $\Delta$ can degenerate into a line segment when some pair of the vertices coincide. We define the length of the perimeter of $\Delta$ as

$$l(\Delta) = |p_1 - p_2| + |p_2 - p_3| + |p_3 - p_1|.$$

For a set $X$ in the plane and $L \geqslant 0$, let

$$\mathscr{T}_X = \{\Delta \text{ a triangle} \mid \text{there is a point } p \text{ such that } p + \Delta \subset X\},$$
$$\mathscr{L}_L = \{\Delta \text{ a triangle} \mid l(\Delta) \leqslant L\}.$$

int $X$ denotes the topological interior of the set $X$, and similar to what was done with translates we let

$$\alpha X = \{\alpha x \mid x \in X\}.$$

For sets $A$, $B$ we denote the difference set

$$A \setminus B = \{a \mid a \in A \text{ and } a \notin B\}.$$

$\mathbb{R}^2$ denotes the Euclidean plane.

We now wish to look at the covering problem from another point of view. We investigate the triangles that have no translate in the interior of $X$ and relate the minimum of their lengths to the maximum length that insures that a triangle (or any closed curve) with that length can be translated into $X$.

LEMMA 4. *Let $X$ be a compact convex set with non-empty interior in the plane. Then there is a triangle (possibly degenerating into a line segment) $\triangle_0 \in \mathcal{T}_X \setminus \mathcal{T}_{\text{int } X}$ such that*

$$l(\triangle_0) = \inf\{ l(\triangle) | \triangle \text{ is a triangle with } \triangle \notin \mathcal{T}_{\text{int } X} \}$$
$$= \sup\{ L | \mathcal{L}_L \subset \mathcal{T}_X \}.$$

*Proof.* Since triangles with sufficiently small perimeter clearly can be translated into $X$, for some $\varepsilon > 0$, $\mathcal{L}_\varepsilon \subset \mathcal{T}_X$. Thus $\sup\{ L | \mathcal{L}_L \subset X\} = L_0$ exists. Let $\triangle \in \mathcal{L}_{L_0}$ with $l(\triangle) = L_0$. Then for every $0 < \alpha < 1$, $\alpha\triangle$ has a translate in $X$, since $l(\alpha\triangle) = \alpha L_0 < L_0$. Thus $\alpha\triangle \in \mathcal{L}_{\alpha L_0} \subset \mathcal{T}_X$. As $\alpha \to 1$, a limit of translates of $\alpha\triangle$ will be contained in $X$, and since $X$ is compact, $\triangle \in \mathcal{T}_X$. Thus $\mathcal{L}_{L_0} \subset \mathcal{T}_X$.

For every $\delta > 0$, let $\triangle_\delta$ be a triangle with $\triangle_\delta \notin \mathcal{T}_X$ with $L_0 < l(\triangle_\delta) \leqslant L_0 + \delta$. By fixing a vertex of $\triangle_\delta$ and bounding all the $\triangle_\delta$, we may choose $\triangle_\delta$ such that $\lim_{\delta \to 0} \triangle_\delta = \triangle_0$, a triangle in the plane. Then by the continuity of $l$, $l(\triangle_0) = L_0 = \sup\{ L | \mathcal{L}_L \subset \mathcal{T}_X\}$, and $\triangle_0 \in \mathcal{T}_X$.

If a triangle $\triangle$ has $l(\triangle) < L_0$, then for some $\alpha > 1$, $l(\alpha\triangle) = \alpha l(\triangle) < L_0$, $\alpha\triangle \in \mathcal{T}_X$, and $\triangle \in \mathcal{T}_{\text{int } X}$. Thus

$$L_0 \leqslant \inf\{ l(\triangle) | \triangle \notin \mathcal{T}_{\text{int } X} \}.$$

Similarly for every $\delta > 0$, for $\triangle_\delta$ chosen as above, $\triangle_\delta \notin \mathcal{T}_{\text{int } X}$ implies that $L_0 = l(\triangle_0) \geqslant \inf\{ l(\triangle) | \triangle \notin \mathcal{T}_{\text{int } X}\}$. Thus $\sup\{ L | \mathcal{L}_L \subset \mathcal{T}_X\} = L_0 = \inf\{ l(\triangle) | \triangle \notin \mathcal{T}_{\text{int } X}\}$.

Since $\triangle_0 = \lim_{\delta \to 0} \triangle_\delta$ and $\triangle_\delta \notin \mathcal{T}_{\text{int } X}$, then $\triangle_0 \notin \mathcal{T}_{\text{int } X}$. Thus $\triangle_0 \in \mathcal{T}_X \setminus \mathcal{T}_{\text{int } X}$. □

The following Lemma is a version of Sperner's Lemma for open (convex) sets. We show how to deduce this from the statement in Lyusternik [**13**, p. 162], where it is stated for closed sets. The following Lemma is the contrapositive stated for the complementary open sets.

LEMMA 5 (Sperner). *Let $\triangle = \langle p_1, p_2, p_3 \rangle$ be a triangle in the plane with non-empty interior. Let $U_1, U_2, U_3$ be open sets in the plane with the line segment $\langle p_{i-1}, p_{i+1} \rangle \subset U_i$, $i = 1, 2, 3$, (indices taken modulo 3). If $\triangle \subset U_1 \cup U_2 \cup U_3$ then $\triangle \cap U_1 \cap U_2 \cap U_3 \neq \varnothing$.*

*Proof.* Suppose for some $i = 1, 2, 3$, $U_i \cap U_{i+1} \cap \langle p_i, p_{i+1} \rangle \neq \varnothing$. Then $U_1 \cap U_2 \cap U_3 \neq \varnothing$. If $p_i \in U_i$, for some $i = 1, 2, 3$, then $p_i \in U_1 \cap U_2 \cap U_3$.

Let $\triangle \subset U_1 \cup U_2 \cup U_3$, i.e., $\bigcap_{i=1}^3 (\triangle \setminus U_i) = \varnothing$, and from the above we may assume that for all $i = 1, 2, 3$, $p_i \in \triangle \setminus U_i$ and $\langle p_i, p_{i+1} \rangle \subset (\triangle \setminus U_i) \cup (\triangle \setminus U_{i+1})$. Thus by Sperner's Lemma for closed sets $\cup_{i=1}^3 (\triangle \setminus U_i)$ does not contain $\triangle$, i.e., $\triangle \cap U_1 \cap U_2 \cap U_3 \neq \varnothing$, finishing the Lemma. □

Let $H = \langle H_1, H_2, H_3 \rangle$ be three closed half planes. We say that $H$ is *nearly bounded* if $H_1 \cap H_2 \cap H_3 = \bar{\triangle}$ is contained between two parallel lines. In the case when two of the $H_i$'s are the same we can write $H = \langle H_1, H_2 \rangle$. If $\bar{\triangle}$ is bounded, then $H$ is clearly nearly bounded. It is easy to show that the following are equivalent:

1. $H$ is nearly bounded.
2. $\langle p_1 + H_1, p_2 + H_2, p_3 + H_3 \rangle$ is nearly bounded, for all $p_1, p_2, p_3$ in the plane.

3. $(p_1 + H_1) \cap (p_2 + H_2) \cap (p_3 + H_3) = \varnothing$, for some $p_1, p_2, p_3$ in the plane.
4. $\langle(\mathbb{R}^2 \setminus \operatorname{int} H_1), (\mathbb{R}^2 \setminus \operatorname{int} H_2), (\mathbb{R}^2 \setminus \operatorname{int} H_3)\rangle$ is nearly bounded.
5. $\bar{\triangle}$ has no translate contained in $\operatorname{int} \bar{\triangle}$.

If we look at the inward pointing vector $n_i$ perpendicular to the boundary of $H_i$, $i = 1, 2, 3$, we see that $H$ is nearly bounded if and only if the 0 vector is in the convex hull of the $n_i$.

Let $\langle p_1, p_2, p_3\rangle = \triangle$ be a triangle in the plane, and let $H = \langle H_1, H_2, H_3\rangle$ be closed half planes with $\bar{\triangle} = H_1 \cap H_2 \cap H_3$. If $H$ is nearly bounded and $p_i \notin \operatorname{int} H_i$, for $i = 1, 2, 3$, then we say that $H$ (or $\bar{\triangle}$) *separates* $\triangle$ (from any set contained in $\bar{\triangle}$). See Figure 4.

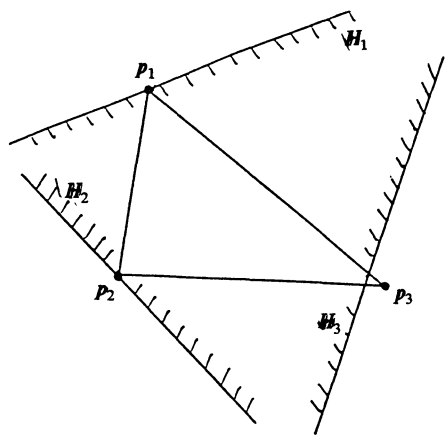



**Fig. 4**

We can now provide a criterion for when a triangle cannot be translated into a convex set.

LEMMA 6. *Let $X$ be a compact convex set with non-empty interior. A triangle (possibly degenerating into a line segment) $\triangle \notin \mathscr{T}_{int\,X}$ if and only if there are closed (nearly bounded) half planes $H = \langle H_1, H_2, H_3\rangle$ that separate a translate of $\triangle$ from $X$, or $H = \langle H_1, H_2\rangle$ (with parallel boundaries) separates a translate of an edge of $\triangle$ from $X$.*

*Proof.* Suppose $H = \langle H_1, H_2, H_3\rangle$, which is nearly bounded, separates $\triangle = \langle p_1, p_2, p_3\rangle$ with $H_1 \cap H_2 \cap H_3 = \bar{\triangle} \supset X$. If a translate of $\triangle$, $p + \triangle \subset \operatorname{int} X$, $p + H_i \subset \operatorname{int} H_i$, $i = 1, 2, 3$, and $p + \bar{\triangle} \subset \operatorname{int} \bar{\triangle}$, which is impossible by property 5. Thus $\triangle \notin \mathscr{T}_{\operatorname{int} X}$. If $H = \langle H_1, H_2\rangle$ separates $\langle p_1, p_2\rangle$, the proof is similar.

Suppose $\triangle \notin \mathscr{T}_{\operatorname{int} X}$. By Lemma 2, $\bigcap_{i=1}^{3}(-p_i + \operatorname{int} X) = \varnothing$.

If for some $i$, $(-p_{i-1} + \operatorname{int} X) \cap (-p_{i+1} + \operatorname{int} X) = \varnothing$ (indices taken mod 3), then a line $B$ separates $(-p_{i-1} + \operatorname{int} X)$ and $(-p_{i+1} + \operatorname{int} X)$ in the usual sense. Thus there is a point $p$ on $B$ and closed half planes $H_{i-1}$ and $H_{i+1}$ such that $p \in bdy\, H_{i-1} \cap bdy\, H_{i+1} = B$, $-p_{i-1} + \operatorname{int} X \subset H_{i-1}$, and $-p_{i+1} + \operatorname{int} X \subset H_{i+1}$, where *bdy* denotes the topological boundary. Thus $p_{i-1} + p \in bdy(p_{i-1} + H_{i-1})$, $p_{i+1} + p \in bdy(p_{i+1} + H_{i+1})$, $p_{i-1} + H_{i-1} \supset X$, $p_{i+1} + H_{i+1} \supset X$, and thus $\langle p_{i-1} + H_{i-1}, p_{i+1} + H_{i+1}\rangle$ is nearly bounded and separates $\langle p + p_{i-1}, p + p_{i+1}\rangle$, an edge of $p + \triangle$, from $X$.

So we are left with the case when $(-p_{i-1} + \text{int } X) \cap (-p_{i+1} + \text{int } X) \neq \varnothing$, for all $i = 1,2,3$. Let $q_i \in (-p_{i-1} + \text{int } X) \cap (-p_{i+1} + \text{int } X)$, for $i = 1,2,3$. We choose $\triangle' = \langle q_1, q_2, q_3 \rangle$ with non-empty interior as a triangle. Then $\langle q_{i-1}, q_{i+1} \rangle \subset -p_i + \text{int } X$, for all $i = 1,2,3$. Thus Lemma 5 and Lemma 2 imply that $\bigcup_{i=1}^{3}(-p_i + \text{int } X)$ does not contain $\triangle'$. Let $q \in \triangle' \setminus \bigcup_{i=1}^{3}(-p_i + \text{int } X)$. Then for $i = 1,2,3$, let $H_i$ be a closed half plane with $q \in bdy\, H_i$ and $(-p_i + \text{int } X) \subset \text{int } H_i$. Thus $q \in \bigcap_{i=1}^{3}(\mathbb{R}^2 \setminus \text{int } H_i) = \triangle''$, and $\triangle'' \cap \langle q_{i-1}, q_{i+1} \rangle = \varnothing$, for $i = 1,2,3$, because $H_i \supset \langle q_{i-1}, q_{i+1} \rangle$. Thus $\triangle'' \cap bdy\,\triangle' \neq \varnothing$ and $\triangle'' \subset \triangle'$. Thus $\triangle''$ is bounded. (It turns out that $\triangle'' = \{q\}$.) Thus $H = \langle H_1, H_2, H_3 \rangle$ is nearly bounded by property 4. $p_i + q \in bdy(p_i + H_i)$ and so $\langle p_1 + H_1, p_2 + H_2, p_3 + H_3 \rangle$ is (nearly) bounded by property 2 and separates $q + \triangle = \langle q + p_1, q + p_2, q + p_3 \rangle$ from $X$. See Figure 5. $\square$

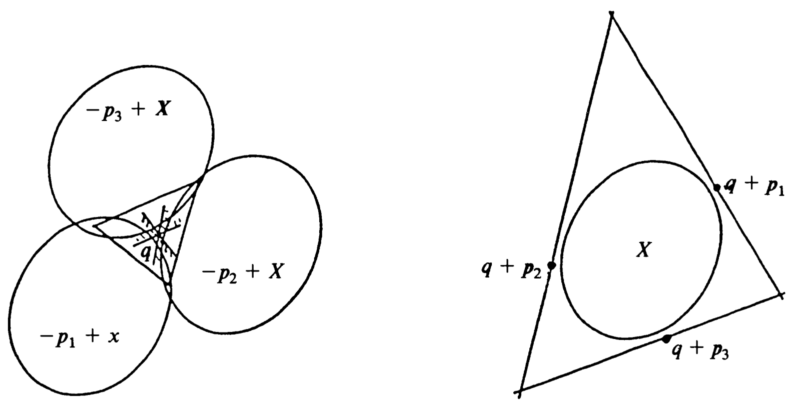



FIG. 5

*Remark 2.* All of the above Lemmas and their proofs generalize to Euclidean space of any (positive) dimension. Both Helly's Theorem and Sperner's Lemma are true and well known in this generality.

*Remark 3.* We can use other functions for $l$ besides the length of the perimeter. The properties of $l$ that we need are that $l$ is a real-valued function defined for at least convex sets in Euclidean space and the following hold.

1. $l$ is continuous.
2. $l(\triangle) < l(\alpha\triangle)$, for $\alpha > 1$.
3. $l(p + \triangle) = l(\triangle)$ for all points $p$.
4. $\{\triangle \mid l(\triangle) \leqslant 1, 0 \in \triangle\}$ is bounded.

For instance we could take

$$l(X) = \text{diameter of } X = \sup\{|x - y| \mid x, y \in X\}.$$

**III. Billiard triangles.** We can now further reduce the set of triangles that need to be considered for our covering problem. Let $X$ be a compact convex set with non-empty interior in the plane. For $i = 1,2,3$, let $p_i$ be three distinct points on the boundary of $X$, and let $H_i$ be a closed half plane containing $X$ with $p_i \in bdy\, H_i$. If

$p_{i+1} - p_i$ and $p_{i-1} - p_i$ (regarded as vectors) make equal angles with the line *bdy* $H_i$ at $p_i$, for all $i = 1, 2, 3$ (indices taken modulo 3), then we say $\triangle = \langle p_1, p_2, p_3 \rangle$ is a *billiard triangle* for $X$. Note that the $H_i$ containing $p_i$ as above, in general, may not be unique. However, we will only consider billiard triangles that are separated from $X$ by a unique $\langle H_1, H_2, H_3 \rangle$, where $\bar{\triangle} = H_1 \cap H_2 \cap H_3$ is a triangle. See Figure 6 below.

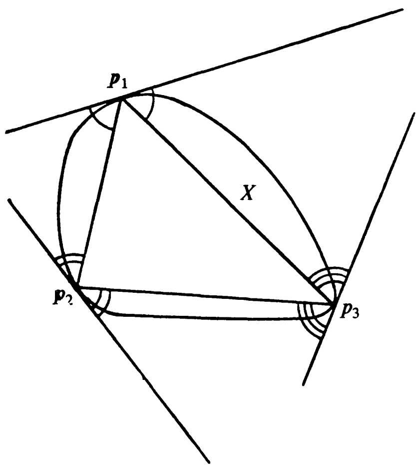



FIG. 6

We look at our covering problem for the special case when $X$ is a triangle. This is a very old problem of Fagnano with some pleasant solutions involving a "reflection principle." Some very elementary proofs are described in Rademacher and Toeplitz [**16**, p. 27–34], as well as Kazarinoff [**10**, pp. 75–77].

LEMMA 7 (Fagnano). *Let $\bar{\triangle}$ be a triangle in the plane. If $\bar{\triangle}$ is acute then the triangle $\triangle$ with smallest perimeter inscribed in $\bar{\triangle}$ is the pedal triangle formed by the feet of the three altitudes of $\bar{\triangle}$. Furthermore $\triangle$ is the unique billiard triangle for $\bar{\triangle}$. If $\bar{\triangle}$ is not acute, then there is no billiard triangle in $\bar{\triangle}$, and any inscribed triangle has perimeter greater than twice the minimum altitude.*

We state the main criterion by which we can calculate which closed curves can be covered by a given compact convex set $X$.

Let $m_\triangle = \inf\{l(\triangle) | \triangle \text{ is a billiard triangle for } X\}$. If $X$ has no billiard triangle $m_\triangle = \infty$. Let $b$ be the minimum breadth of the set $X$, i.e., $b$ is the minimum distance between two distinct parallel support lines of $X$.

LEMMA 8. *Let $X$ be a compact convex set. Any closed curve of length one or less can be covered by a translate of $X$ if and only if* $\min\{m_\triangle, 2b\} \geqslant 1$.

*Proof.* We will show that $\min\{m_\triangle, 2b\} = l(\triangle_0)$, where $\triangle_0 \in \mathscr{T}_X \setminus \mathscr{T}_{\text{int}\, X}$ is as in Lemma 4. The result will then follow using Lemma 3 as well as Lemma 4.

Suppose that $\triangle_0$ does not degenerate to a line segment. If $\langle H_1, H_2 \rangle$ separates an edge of $\triangle_0$ from $X$, then $\triangle_0$ degenerates to a line segment. Thus by Lemma 6 there

are half planes $H = \langle H_1, H_2, H_3 \rangle$ that separate a translate of $\triangle_0$ from $X$. We take $\triangle_0 \subset X$. By Lemma 7 (Fagnano) the pedal triangle of $\bar{\triangle} = H_1 \cap H_2 \cap H_3$ is the shortest triangle separated by (i.e. inscribed in) $\bar{\triangle}$, if there is such a shortest triangle. Thus $\triangle_0$ must be this pedal triangle for $\bar{\triangle}$. So $\triangle_0$ has the billiard property at each $p_i$, $i = 1, 2, 3$, $\bar{\triangle}$ is unique, and thus $l(\triangle_0) \geqslant m_\triangle$. By Lemma 4, Lemma 6, and the definition of a billiard triangle $l(\triangle_0) \leqslant m_\triangle$. Thus $l(\triangle_0) = m_\triangle \leqslant 2b$.

If $\triangle_0$ is a line segment, then clearly $l(\triangle_0) = 2b \leqslant m_\triangle$. □

*Remark* 4. If we replace $l$ (the function that gives the length of the perimeter) by the diameter function, we can also provide a similar analysis as above. The critical triangle $\triangle_0$ of Lemma 4 can be taken to be either an equilateral triangle, an isosceles triangle with its base shorter than its two legs, or a line segment, since any such $\triangle_0$ is contained in one such triangle.

A corollary of the above observation is that if every circular wedge with a 60° vertex angle and leg length one can be translated into $X$, a compact convex set, then any set of diameter one or less can be translated into $X$. Of course the same is true for the Reuleaux triangle (see the next section) of diameter one, replacing the wedge. This is a curious dual to Chakerian's Theorem mentioned in section V.

We can continue with an analogue of Lemma 8. Namely $\triangle_0$ must be separated by $\bar{\triangle}$ in only a particular way.

If $\triangle_0$ is an equilateral triangle, then the lines perpendicular to the edges of $\bar{\triangle}$ at $p_i$, $i = 1, 2, 3$, must meet at a point in $\triangle_0$. See Figure 7. If the perpendiculars do not meet at a point as above, $\triangle$ can be "rotated" out of $\bar{\triangle}$ and then contracted to obtain another $\triangle'$, still separated by $\bar{\triangle}$ but with $l(\triangle') < l(\triangle)$ ($l$ now being the diameter).

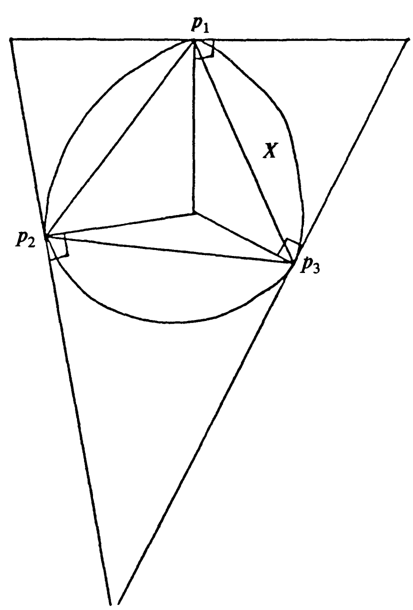



FIG. 7

If $\triangle_0$ is an isosceles triangle, then the equal sides of $\triangle_0$ must be perpendicular to the corresponding sides of $\bar{\triangle}$, as in Figure 8 below.

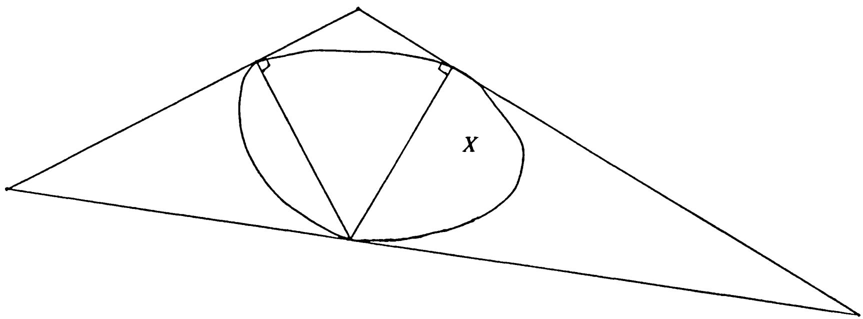



FIG. 8

If $\triangle_0$ is a line segment, then $\triangle_0$ is of minimal breadth, as before.

**IV. Reuleaux triangles.** We now apply Lemma 8 to prove a special case of the Theorem, namely when $X$ is a Reuleaux triangle. A *Reuleaux triangle* is the convex set of constant breadth $r$ obtained as the intersection three circular disks of the same radius $r$ with the centers of the circles at the vertices of an equilateral triangle of side length $r$. See Figure 9.

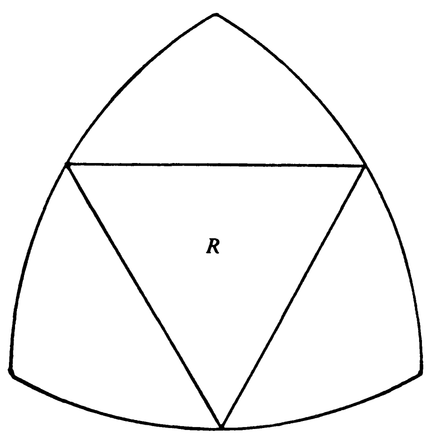



FIG. 9

LEMMA 9. *Let $R$ be a Reuleaux triangle of constant breadth $1/2$. Then $R$ has only the one symmetric billiard triangle joining the midpoints of the circular arcs as below. Thus $m_{\triangle} = 3(\sqrt{3} - 1)/2 \approx 1.098 > 1$ and every closed curve of length one or less is covered by a translate of $R$.*

*Proof.* From Lemma 7 the only billiard triangle in a triangle $\bar{\triangle}$ is the pedal triangle $P$ formed by the feet of the three altitudes of $\bar{\triangle}$, and this billiard triangle exists only when $\bar{\triangle}$ is acute.

Suppose $\triangle = \{p_1, p_2, p_3\}$ forms a billiard triangle in $R$. Let $L_i$, $i = 1, 2, 3$, be the corresponding support lines. Clearly each $p_i$ is on a separate circular arc of the Reuleaux triangle $R$. Let $r_i$ be the vertex of $R$ opposite $p_i$, and let $C$ be the center of the inscribed circle of $\triangle$. Then $r_i$, $C$, $p_i$, are collinear because $r_i - p_i$ and $C - p_i$ are both perpendicular to $L_i$. See Figure 10.

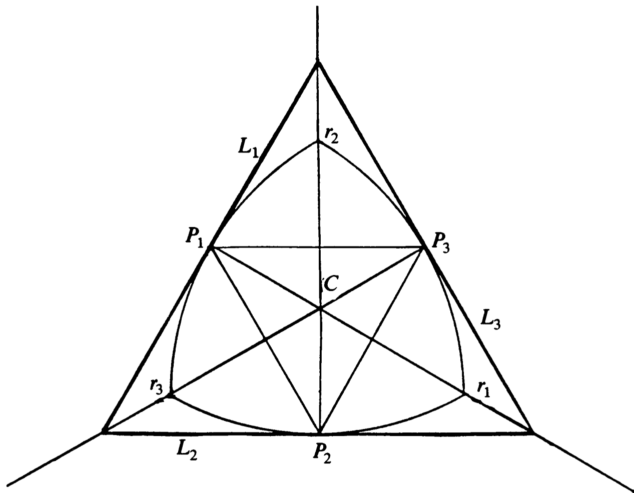



FIG. 10

Thus $\triangle$ is a billiard triangle for the triangle $\overline{\triangle}$ formed by $L_1$, $L_2$, $L_3$, as well as $R$. Thus $\triangle$ is the pedal triangle of $\overline{\triangle}$, and $C$ is the altitude center of $\overline{\triangle}$. With all this information we wish to show that $C$ is the center of $R$. This will show that $\triangle$ is the desired symmetric triangle.

Let $C_R$ be the center of $R$. Let $q_i$ be the midpoint of the circular arc from $r_{i+1}$ to $r_{i-1}$, as in Figure 11 below.

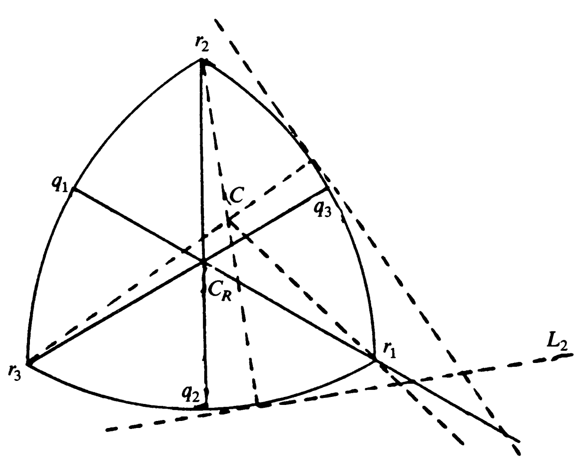



FIG. 11

Assume, without loss of generality, that $C$ lies in the closed region of $R$ defined by $C_R$, $q_3$, $r_2$. It is clear that the support lines $L_3$ and $L_2$ intersect at a point on the $q_3$ side of the line $C_R r_1$. On the other hand the ray from $r_1$ outside $R$ on the line $Cr_1$ lies on the other side of $C_R r_1$. Thus the three lines $L_2, L_3, Cr_1$ intersect only if $C = C_R$ and $p_i = q_i$, $i = 1, 2, 3$. □

**V. Proof of the theorem.** We thank one of the referees for pointing out that, at this point, we can complete the proof of the Theorem by using the following result of Chakerian [**3**, p. 760].

THEOREM (Chakerian). *Let $R$ be a Reuleaux triangle of constant breadth $b$ in the Euclidean plane, and let $P$ be a compact subset of the plane. Suppose that each congruent copy of $P$ can be covered by a translate of $R$. Then if $K$ is any compact convex set of constant breadth $b$, each congruent copy of $P$ can be covered by a translate of $K$.*

*Remark* 5. The proof of Chakerian's Theorem above uses yet another variation of Helly's Theorem and a simple property of sets of constant breadth, and it is not difficult. However, we include another proof using the ideas of finding billiard triangles using Lemma 8 above.

By Lemma 8 we need only prove that all billiard triangles $\triangle = \langle p_1, p_2, p_3 \rangle \subset X$ have $l(\triangle) \geqslant 1$, since we assume $b = 1/2$. Thus we suppose $\triangle$ is a billiard triangle with $l(\triangle) < 1$, and we look for a contradiction.

Let $q_1, q_2, q_3$ be the points on the boundary of $X$ such that the segment $\langle p_i, q_i \rangle$ is perpendicular to the support line for $X$ at $p_i$, $i = 1, 2, 3$. Note that the segments are coincident at the incenter of $\triangle$, each $q_i$ is outside $\triangle$ since $q_i$ is on the boundary of $X$, and $|p_i - q_i| = 1/2$ since $X$ is of constant breadth $1/2$.

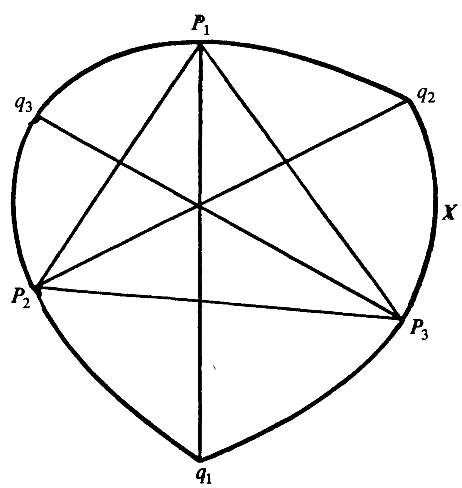



FIG. 12

Consider for $t \geqslant 1$, $r_i(t) = tq_i + (1-t)p_i$. Let

$$S_t = \{p_1, p_2, p_3, r_1(t), r_2(t), r_3(t)\},$$
$$t_0 = \sup\{t \mid \text{diameter of } S_t \leqslant t/2\}.$$

Note that $|r_i(t) - p_i| = t/2$. The supremum exists since the diameter of $S_1 = 1/2$, and for some $i$ and $j$ the angle between $q_i - p_i$ and $q_j - p_j$ is greater than $60°$ and thus

$$\lim_{t\to\infty} |r_i(t) - r_j(t)|/t > 1.$$

Let $r_i = r_i(t_0)$, $i = 1,2,3$. Then for some $i \neq j$ either $|p_i - r_j| = t_0/2 =$ diameter $S_{t_0}$ or $|r_i - r_j| = t_0/2$. By renumbering, one of the following cases occurs.

*Case* 1. $|p_1 - r_2| = t_0/2$.
We then have the following diagram where •——• indicates that the distance is $t_0/2$.

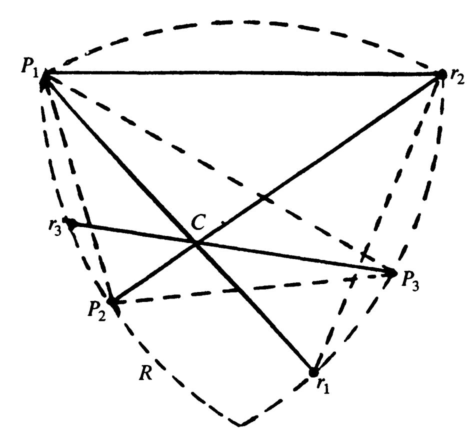



FIG. 13

Let $R$ be the Reuleaux triangle with $p_1$ and $r_2$ as vertices with the third vertex on the same side of $[p_1, r_2]$ as $p_2$ and $r_1$. Thus the breadth of $R$ is $t_0/2$. Then $p_2, r_1$ are on the boundary of $R$ on the opposite sides of $r_2, p_1$ respectively, since $|p_1 - r_1| = |p_2 - r_2| = t_0/2$, $|p_1 - p_2| \leqslant t_0/2$, and $|r_1 - r_2| \leqslant t_0/2$. Thus $\langle p_2, r_2\rangle \cap \langle p_1, r_1\rangle = C$ is a point in the interior of $R$. Thus $r_3$ and $p_3$ lie in the open cones determined by $Cp_2, Cp_1$ and $Cr_2, Cr_1$ respectively. Thus $|r_3 - r_2| \leqslant t_0/2$, $|r_3 - p_1| \leqslant t_0/2$ implies $r_3$ is in $R$ and $|p_3 - p_1| \leqslant t_0/2$, $|p_3 - r_2| \leqslant t_0/2$ implies $p_3$ is in $R$ as well. The diameter of $R$ is $t_0/2$. Thus $r_3$ and $p_3$ must be on the boundary with at least one of them equal to $p_1$ or $r_2$, which is impossible. Thus Case 1 cannot occur.

*Case* 2. $|r_1 - r_2| = t_0/2$.
We then have the following diagram, Figure 14.

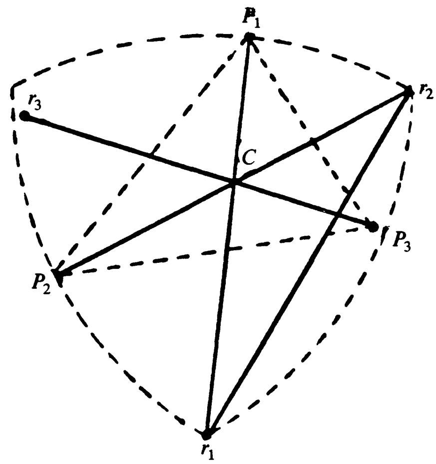



FIG. 14

Let $R$ be the Reuleaux triangle with $r_1, r_2$ as vertices as before and thus with breadth $t_0/2$ again. As in Case 1 $\langle p_1, r_1\rangle \cap \langle p_2, r_2\rangle = C$, a point in the interior of $R$ on the same side of $\langle r_1, r_2\rangle$ as $p_1, p_2, r_3$. Since $|r_3 - r_2| \leqslant t_0/2$ and $|r_3 - r_1| \leqslant t_0/2$, $r_3$ must be in $R$. Since $|r_3 - p_3| = t_0/2$, either $p_3$ is outside of $R$ altogether, or $p_3$ is on the arc of $R$ from $r_1$ to $r_2$ with $r_3$ as the opposite vertex of $R$. In either case let $\bar{p}_3$ be the point on the circular arc from $r_1$ to $r_2$ of $R$ on $[r_3, p_3]$. This point exists since $Cp_3$ is in the cone determined by $Cr_2$ and $Cr_1$.

Let $H_1, H_2, H_3$ be the support half planes for $R$ at $p_1, p_2, \bar{p}_3$ respectively. Clearly $H_1 \cap H_2 \cap H_3$ is bounded and thus by Lemma 6, $\triangle$ cannot be covered by int $R$. Thus by Lemma 9

$$1 > l(\triangle) \geqslant t_0 \geqslant 1,$$

the final contradiction.

To finish the Theorem suppose any closed curve of length one or less is covered by a translate of $X$. Then the minimum breadth of $X$ is $b \geqslant 1/2$ by Lemma 8. Let $b(\theta)$ be the breadth in direction $\theta$. Then it is a well-known formula of Cauchy, see Hurwitz [8] for example (or Chakerian and Klamkin [4] for a similar use), that the length of the perimeter of $X$ is

$$l(X) = (1/2)\int_0^{2\pi} b(\theta)\, d\theta \geqslant (1/2)2\pi(1/2) = \pi/2,$$

with equality if and only if $b(\theta) = 1/2$, a constant. $\square$

**VI. Wetzel's problem.** We have found all the sets of shortest perimeter whose translates cover any closed curve of length one or less. As mentioned in the introduction, Wetzel considered the question: What is a set of smallest area whose translates cover any closed curve of length one or less? In other words, what is a translation cover of smallest area for $\mathscr{L}_1$ (the collection of closed curves of length one or less)?

Among all compact convex sets of constant breadth $1/2$, the Reuleaux triangle is the set of smallest area $(\pi - \sqrt{3})/8 \approx 0.17619$, and of course is a translation cover

for $\mathscr{L}_1$. Wetzel [21] found a set shown in Figure 15 (with some of the billiard triangles indicated), which is also a translation cover for $\mathscr{L}_1$ and has area $\approx 0.17141$ (and thus is necessarily not of constant breadth). (Wetzel claimed incorrectly that the area is $\approx 0.15900$.)

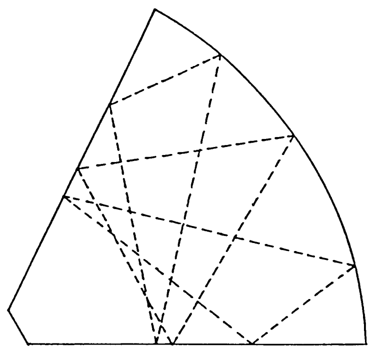


FIG. 15

Figure 16 shows a set, looking a bit like a church window, with area $1/6 \approx 0.16667$ which is also a translation cover for $\mathscr{L}_1$. The base has length $1/2$ and the height is also $1/2$. The curves are parabolas with their line of symmetry about the base. Since the common focus of the two parabolas is the midpoint of the base, it is easy to show that the indicated isosceles triangles (with horizontal base and vertex at the midpoint of the base of the church window) are indeed billiard triangles. It is a bit more difficult to show that there are no more shorter billiard triangles.

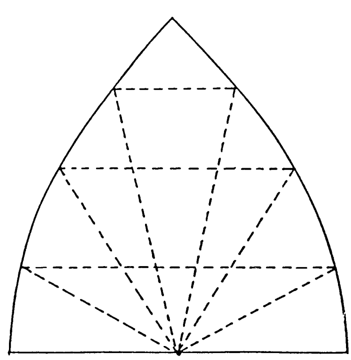


FIG. 16

We do not know what the minimum area is for translation covers for $\mathscr{L}_1$, but there are still other translation covers for $\mathscr{L}_1$ that are more complicated but have area smaller than $1/6$, the area of the church window. The best translation cover that we have found has area $\approx 0.16526$. However, Wetzel [21], modifying an argument of Pal [15], derived a lower bound $\approx 0.15544$ for the area of any translation cover for $\mathscr{L}_1$.

Note that the Reuleaux triangle, Figure 9, and the church window, Figure 16, are *minimal* with respect to being a translation cover for $\mathscr{L}_1$, i.e., they do not contain a proper closed convex subset that is also a translation cover for $\mathscr{L}_1$. A compact convex set $X$ is a minimal translation cover for $\mathscr{L}_1$ if and only if it is the convex hull of all its billiard triangles of length 1, including the degenerate case of a doubly covered minimum breadth of length $1/2$. Wetzel's wedge, Figure 15, is not minimal since a small amount of the set, near the lower left-hand corner, can be shaved off, and it will still remain a translation cover for $\mathscr{L}_1$.

**VII. Related problems.** As mentioned in Remark 3 and Remark 4, it is possible to use a similar analysis as in section II to consider other classes of objects to be covered, as well as minimizing with respect to functions other than the length of the perimeter. We mention, without proof, some results for some of these combinations.

We consider three classes of objects, in the plane, to be covered:

1. $\mathscr{L}_1$, closed curves of length one or less.
2. Compact sets of diameter one or less.
3. Arcs of length one or less.

Then we consider translation covers for each of the classes above and minimize with respect to the following:

1. Length of the perimeter.
2. Diameter, the largest distance between pairs of points in the set.
3. Area.

The results are summarized in the following table.

| | minimizing functions for the translation cover | | |
|---|---|---|---|
| sets to be covered | length of perimeter | diameter | area |
| closed curves of length $\leqslant 1$ | all sets of constant breadth $1/2$, length $= \pi/2$ | all sets of constant breadth $1/2$, diameter $= 1/2$ | ? $.155 <$ area $< .167$ |
| sets of diameter $\leqslant 1$ | a circle of diameter $2\sqrt{3}/3$, length $= 2\pi\sqrt{3}/3$ | a circle of diameter $2\sqrt{3}/3$, (Jung's Thm. [**9**]) | ? |
| arcs of length $\leqslant 1$ | all sets of constant breadth 1, length $= \pi$ | all sets of constant breadth 1, diameter $= 1$ | the equilateral triangle of side length $2\sqrt{3}/3$ (Pal's Thm. [**15**]) area $= \sqrt{3}/3$ |

Each entry in the table represents the whole collection of minimizing sets, if known. A ? indicates that the minimizing object is unknown to us.

The line corresponding to closed curves of length $\leqslant 1$ has already been discussed, except for the entry under diameter. However, from our main Theorem, any set of constant breadth $1/2$ forms a translation cover for all sets of perimeter of length $\leqslant 1$. In order to cover any doubly covered line segment, any translation cover $X$ for $\mathscr{L}_1$ must have minimum breadth $\geqslant 1/2$, and thus $X$ must have diameter $\geqslant 1/2$. If $X$ has diameter $= 1/2$ and minimum breadth $\geqslant 1/2$, then $X$ must be a set of constant breadth $1/2$.

For the line corresponding to the sets of diameter $\leqslant 1$, Jung's Theorem [**9**] implies that the circle of diameter $2\sqrt{3}/3$ is a translation cover for such sets, but it does not imply, and it is nontrivial to prove, that, in the case of the perimeter and the diameter, this circle is the only such minimal convex set.

For the line corresponding to arcs of length $\leqslant 1$, it turns out that $X$ is a translation cover for this collection if and only if $X$ has minimum breadth 1. A theorem of Pal [**15**] states that the equilateral triangle has the least area among all sets of minimum breadth 1.

In dimensions greater than two we have very little information. For a given compact convex set $X$ in $\mathbb{R}^d$, if one can calculate the length $L$ of the shortest billiard path with $d+1$ or fewer bounces, $X$ is a translation cover for curves of length $L$ or less. For example, in *Martin Gardner's Sixth Book of Mathematical Diversions* [**7**, pp. 29–38], it is stated that the regular tetrahedron, with side length one, has three billiard paths, each of total length $4/\sqrt{10}$, with four bounces. It turns out that these billiard paths are the shortest, and thus the regular tetrahedron of side length $\sqrt{10}/4$ is a translation cover for closed curves of length one or less. Similarly the regular simplex of side length one in $\mathbb{R}^d$ has $d!/2$ billiard paths of minimal length $(d+1)\sqrt{6/d(d+1)(d+2)}$; and thus the regular simplex of side length $\sqrt{d(d+1)(d+2)/6}/(d+1)$ is a translation cover for closed curves of length one or less in $\mathbb{R}^d$.

REFERENCES

1. E. Barbier, Note sur le probleme de l'aiguille et le jeu joint couvert, *J. Math. Pures Appl. Ser. 2*, 5(1860) 273–286.
2. T. Bonnesen and W. Fenchel, Theorie der konvexen Korper. Springer Verlag, Berlin, 1974.
3. G. D. Chakerian, Intersection and covering properties of convex sets, this MONTHLY, 76(1969) 753–766.
4. G. D. Chakerian and M. S. Klamkin, Minimal covers for closed curves, *Math. Mag.*, 46(1973) 55–61.
5. L. Danzer, B. Grünbaum, and V. Klee, Helly's theorem and its relatives, AMS, *Proc. Sympos. Pure Math.*, 7(1963) 101–180.
6. H. G. Eggleston, Convexity, Cambridge, 1969.
7. M. Gardner, *Martin Gardner's Sixth Book of Mathematical Diversions from the Scientific American*, University of Chicago Press.
8. A. Hurwitz, Sur quelques applications geometriques des series de Fourier, *Annales de L'Ecole Normale Superieure*, 3me serie 19, 357–408.
9. H. W. E. Jung, Über die kleinste Kegel, die eine räumliche Figur einschliesst, *J. Reine Angew. Math.*, 123 (1901) 241–257.
10. N. D. Kazarinoff, Geometric Inequalities, Random House, 1961.
11. V. Klee, The critical set of a convex body, *Amer. J. of Math.*, 75(1953) 178–188.
12. S. R. Lay, Convex Sets and their Applications, Wiley, 1982.
13. L. A. Lyusternik, Convex Figures and Polyhedra, Dover, 1963.
14. J. C. C. Nitche, The smallest sphere containing a rectifiable curve, this MONTHLY, 78(1971) 881–882.
15. J. Pal, Ein Minimumproblem fur Ovale, *Math. Ann.*, 83(1921) 311–319.
16. H. Rademacher and O. Toeplitz, The Enjoyment of Mathematics, Princeton University Press, 1957.
17. J. Schaer and J. E. Wetzel, Boxes for curves of constant length, *Israel J. Math.*, 12(1972) 257–265.
18. J. E. Wetzel, Triangular covers for closed curves of constant length, *Elemente der Mathematik*, 25/4 (1970) 78–81.
19. ______, Covering balls for curves of constant length, *L'Enseignement Math.*, 17(1971) 275–277.
20. ______, On Moser's problem of accommodating closed curves in triangles, *Elemente der Mathematik*, 27/2 (1972) 35–36.
21. ______, Sectorial covers for curves of constant length, *Canad. Math. Bull.*, Vol 16 (3) (1973) 367–375.

# Reconstructing a Function from Its Set of Tangent Lines

ALAN HORWITZ, *Pennsylvania State University, Media, PA 19063*

ALAN HORWITZ: I received my Ph.D from Temple University in 1984 in complex variables and approximation theory under the direction of Donald Newman. My research interests have included polynomial interpolation, Banach spaces of analytic functions, and algebraic and analytic differential equations. The "inverse-type problem" considered in this article is related to work I did with Lee Rubel in polynomial interpolation.

**1. Introduction.** Given a differentiable function $f$ defined on an interval $I = [a, b]$, one can easily find its set of tangent lines $T(f)$ relative to $I$. $L$ is in $T(f)$ if and only if $L$ is given by the equation

$$y = f(x_0) - x_0 f'(x_0) + x f'(x_0) \quad \text{for some } x_0 \text{ in } I.$$

Can we reverse this process—i.e., given the *set* $T(f)$, how does one find $f$? We stress that we are given $T(f)$ as a set with no further structure, so that while we know which lines are tangent lines, we do not know *where* they are tangent to the graph of $f$. Of course if we did, then the problem discussed here would be trivial.

Imagine, then, an object traveling on a certain path and giving off signals that enable us to detect the set of tangents to its path without our actually knowing the points of tangency. We then want to reconstruct the path of the object. We will call that the reconstruction problem.

The main results of this paper are contained in Sections 3 and 4, where we solve three related problems. First we solve the reconstruction problem in the following way, where we assume that $f''$ exists and has only finitely many zeros in $I$. Given any tangent line $T$, we simply choose a sequence of tangent lines $\{T_j\}$ converging to $T$, which means convergence of the slopes and $y$-intercepts. Then every subsequence of the sequence of intersection points of $T_j$ and $T$ converges to a point where $T$ is tangent to the graph of $f$. If we do this for each tangent line $T$ we obtain all but a finite number of points on the graph of $f$.

Closely related to the reconstruction problem is the following, which we will call the uniqueness problem. *Must two functions with the same set of tangent lines be identical*? In Theorem 2 we prove that this is indeed the case, again assuming finitely many zeros for the second derivatives of the two functions.

Our interest in the work in this article began, in fact, with the uniqueness problem, which in turn grew out of work with Lee Rubel in Lagrange interpolation (see [2]). It was somewhat surprising that the solution of the uniqueness problem would lead to the solution of another problem of Rubel's in integral geometry. We will call that the cutting line problem, and it goes as follows. Suppose every line in the plane intersects two curves the same number of times, counting multiplicities. Must the two curves be identical? We prove that the answer is yes in Theorem 3 by showing that the two curves must have the same set of tangent lines! Of course we require the same smoothness assumptions as in Theorem 2.

In Section 5 we discuss some connections between the results of Section 3 and the notion of an envelope of a family of curves. In Section 6 we discuss some natural generalizations of our work to higher dimensions, and in Section 7 we show how to recover an analytic function from its set of tangent lines relative to a domain in the complex plane. In Section 8 we give a simple example that shows that our method can fail in general if $f''$ is allowed to have infinitely many zeros in a bounded interval, and we conclude the paper with some open questions.

**2. Preliminaries.** Given two lines $L_1$ and $L_2$ with different slopes, we denote their point of intersection by $i(L_1, L_2) = (i_x(L_1, L_2), i_y(L_1, L_2))$. Denote the $y$-intercept of a line $L$ by $b(L)$ and the slope of $L$ by $m(L)$. We say that a *sequence of lines* $\{L_j\}$ *converges to* $L$ if $b(L_j) \to b(L)$ and $m(L_j) \to m(L)$.

It is easy to prove that if a differentiable function $f$ is, say, convex on $[a, b]$, then $f$ can be recovered from $T(f)$ as follows:

$$f(x_0) = \sup_{T \in T(f)} T(x_0), \quad x_0 \in [a, b].$$

A similar result holds if $f$ is concave on $[a, b]$.

One might try to use this fact in the more general situation where $f''$ is allowed to change sign on $[a, b]$, but this approach is not as easy as it might seem when one is given just the set $T(f)$ and the interval $[a, b]$. Indeed most of the results in this paper are easy to prove if $f''$ never vanishes.

It should also be noted that our approach really does not require knowing the interval in advance, and also works when $T(f)$ is taken relative to a domain in the complex plane where $f$ is analytic and $\sup T(x_0)$ makes no sense.

While it is not hard to extend the results of this paper to curves in general (with similar assumptions on the second derivative), for convenience we will work with graphs of functions.

**3. The reconstruction and uniqueness problems.** Assume throughout this section (unless stated otherwise) that $f$ is a real-valued function such that $f''$ exists and has only finitely many zeros in every bounded interval in its domain (zeros of infinite multiplicity are allowed). Let $T(f)$ be the set of tangent lines to $f$ relative to some interval $I = [a, b]$, and let $T \in T(f)$. Here is how to find the graph of $f$, from $T(f)$.

*Step* 1. Choose any sequence of distinct tangent lines $\{T_j\}$ converging to $T$, with $m(T_j) \neq m(T)$ for all $j$ (it is easily shown that such a sequence always *exists*).

*Step* 2. Find a limit point $P_T$ of the sequence $\{i(T_j, T)\}$.

*Step* 3. Form the set $S = \bigcup P_T$, the union taken over all $T$ in $T(f)$.

It should be noted that one need only choose *some* sequence $\{T_j\}$ converging to $T$ and *some* limit point $P_T$ of $\{i(T_j, T)\}$.

It will follow from Theorem 1 and Proposition 1 that $S$ contains all but a finite number of points on the graph of $f$ (relative to $I$). Since $f$ is continuous we can then easily obtain the remaining points. The reason we might miss any points at all on the graph of $f$ is due to the fact that a line might be tangent at more than one point. We call such a line a *multiple tangent line*. Shortly we shall prove Proposition

1, which states that there are only finitely many such lines, and hence they can be ignored!

We now state and prove the results which justify the claim made above that the set $S$ contains all but finitely many points on the graph of $f$.

THEOREM 1. *Let $\{T_j\}$ be a sequence of distinct tangent lines that converges to a tangent line $T$, $m(T_j) \neq m(T)$ for all $j$. Then the set $V$ of limit points of the sequence $\{i(T_j, T)\}$ is non-empty, and every point of $V$ is a point where $T$ is tangent to the graph of $f$.*

Before proving Theorem 1 we need the following lemma.

LEMMA 1. *Suppose $\{x_j\}$ is a sequence in $I$ that converges to $x_0$, with $T_j$ and $T$ tangent at $(x_j, f(x_j))$ and $(x_0, f(x_0))$ respectively (they may be tangent at other points as well). Assume also that $m(T_j) \neq m(T)$ for all $j$. Then $\{i(T_j, T)\}$ converges to $(x_0, f(x_0))$.*

*Proof.* A simple computation gives:

$$i_x(T_j, T) = x_0 + \frac{f'(x_j) - \dfrac{f(x_j) - f(x_0)}{x_j - x_0}}{\dfrac{f'(x_j) - f'(x_0)}{x_j - x_0}} \equiv x_0 + A_j. \tag{3}$$

We shall show that $A_j \to 0$ (this is trivial if $f''(x_0) \neq 0$ since in that case the denominator in (3) does not tend to 0, while the numerator does). Define

$$B_j = \frac{f'(x_0) - \dfrac{f(x_j) - f(x_0)}{x_j - x_0}}{\dfrac{f'(x_j) - f'(x_0)}{x_j - x_0}} = (x_0 - x_j)\frac{f'(\xi_j) - f'(x_0)}{f'(x_j) - f'(x_0)}, \tag{4}$$

where $\xi_j$ lies between $x_j$ and $x_0$ by the Mean Value Theorem. Now it suffices to show that $B_j \to 0$ since $A_j - B_j = x_j - x_0$ approaches 0. We know that $f'(x)$ is monotonic for $x$ sufficiently close to $x_0$, for otherwise $f''$ would have infinitely many zeros, by Rolle's Theorem. It then follows easily that

$$\left| \frac{f'(\xi_j) - f'(x_0)}{f'(x_j) - f'(x_0)} \right| \leq 1$$

for all $j$ sufficiently large, and thus $B_j \to 0$. Hence $i_x(T_j, T) \to x_0$ which then implies that $i_y(T_j, T) \to f(x_0)$, and the lemma is proved.

*Remark.* The idea of taking intersection points of nearby members of a family of curves has been considered by many authors, and in fact goes back to Leibniz. For example, see [**1**, ch. 5].

*Proof of Theorem* 1. For each $j$, choose a point $(x_j, f(x_j))$ where $T_j$ is tangent to the graph of $f$. Since $\{x_j\} \subseteq I$, there is a subsequence $\{x_{j_k}\}$ that converges to $x_0$ for some $x_0 \in I$. Since $\{T_j\}$ converges to $T$, $T$ must be tangent at $(x_0, f(x_0))$. By

Lemma 1, $\{i(T_{j_k}, T)\}$ converges to $(x_0, f(x_0))$. Thus $V$ is non-empty, and a similar argument completes the proof of the theorem.

As a consequence of Theorem 1 and Lemma 1 we now easily get the following result.

THEOREM 2. *Suppose $f''$ and $g''$ exist and have only finitely many zeros in $I$. Then $T(f) = T(g)$ (as sets) implies $f \equiv g$ on $I$.*

The following Proposition, interesting in its own right, is very useful for getting around the complexities posed by multiple tangent lines. One simply ignores them since there are only finitely many, which simplifies the method for recovering $f$ from $T(f)$, as noted earlier.

PROPOSITION 1. *Suppose $f''$ exists and has only finitely many zeros in $I$. Then there are only finitely many multiple tangent lines.*

*Remark.* We shall use the obvious fact that any given tangent line can be tangent at only finitely many points of the graph of $f$. Note, however, that this is *not* what Proposition 1 says!

*Proof of Proposition* 1. Suppose to the contrary that there is a sequence of distinct multiple tangent lines $\{T_j\}$, say, such that $T_j$ is tangent at both $(x_j, f(x_j))$ and $(t_j, f(t_j))$ with $x_j \neq t_j$ for all $j$. We shall derive a contradiction. By taking subsequences if necessary, assume that $t_j \to t_0$ and $x_j \to x_0$.

*Case* 1. $x_0 = t_0$.
Now $f(x_j) - x_j f'(x_j) = f(t_j) - t_j f'(t_j)$ and $f'(x_j) = f'(t_j)$, which implies

$$\frac{f(x_j) - f(t_j)}{x_j - t_j} = f'(t_j) \quad \text{for all } j. \tag{5}$$

By the Mean Value Theorem this says that $f'(\xi_j) = f'(t_j)$ for $\xi_j$ between $x_j$ and $t_j$. We then have $f'(\xi_j) = f'(t_j) = f'(x_j)$ for all $j$, and this clearly implies that $f''$ has infinitely many zeros since $\{\xi_j\}, \{x_j\}, \{t_j\}$ all converge to $x_0$. (Note that $t_j \neq x_0$, $x_j \neq x_0$ for large $j$ since the $T_j$ are distinct.)

*Case* 2. $x_0 \neq t_0$.
Now $\{T_j\}$ converges to $T$, where $T$ is tangent at $P_0 = (t_0, f(t_0))$ and $Q_0 = (x_0, f(x_0))$. (So $T$ is also a multiple tangent line.) Then by Lemma 1, $\{i(T_j, T)\}$ converges to *both* $P_0$ and $Q_0$. This contradicts the assumption that $x_0 \neq t_0$.

*Remarks.*
1. There are some subtle connections between Lemma 1, Theorem 1, and Proposition 1 which might not be apparent at first glance. We invite the reader to explore some of these connections. For example
Question: To obtain the results of this section, is it sufficient to prove Lemma 1 in the case $f''(x_0) \neq 0$, since $f''$ has only finitely many zeros?
2. As noted earlier, we do not have to be given the interval $I$ along with the set of tangent lines to $f$ in order to recover $f$, as long as we know that $T(f)$ comes from *some bounded* interval. This leaves open the question of recovering $f$ from its set of tangent lines on an unbounded interval, such as the entire real line. While we have

some results in this case, there appear to be difficulties that are not present in the bounded case.

3. If $f$ is *complex-valued*, one can recover its real and imaginary parts by looking at the real and imaginary parts of the elements of $T(f)$.

**4. The cutting line problem.** Let $n_L(f)$ = the number of times the line $L$ intersects the graph of $y = f(x)$, *multiplicities included*. So, for example, if $L$ is the real axis, then $n_L(x^k) = k$ and $n_L(\exp[-1/x^2]) = \infty$. While our result below also follows if multiplicities are not included, it is natural to count a point of contact in the usual way—by the number of derivatives that agree at that point.

THEOREM 3. *Suppose $f''$ and $g''$ exist and have only finitely many zeros in a closed interval $I$, and suppose that $n_L(f) = n_L(g)$ for every line $L$ in the plane. Then $f \equiv g$ on $I$.*

*Proof.* We shall show that $f$ and $g$ have the same set of tangent lines, and the result will then follow immediately from Theorem 2. Consider a set $S$ of tangent lines to $f$ such that every tangent line $T$ in $S$ satisfies:

(i) $T$ is tangent to the graph of $f$ at precisely one point $(c, f(c))$, $c \in I$.

(ii) At the point of tangency of $T, (c, f(c))$, $f''(c) \neq 0$.

(iii) $T$ does not go through an endpoint of the graph of $g$.

It follows that $S$ contains all but *finitely many* of the full set of tangent lines $T(f)$. This is true for (i) by Proposition 1, while it holds for (ii) because $f$ has only finitely many zeros in $I$. For the same reason only finitely many tangent lines can pass through any fixed point $P$ (this is a simple application of Rolle's theorem), and so all but finitely many tangents satisfy (iii). (Assumptions (i) and (ii) in the definition of $S$ are really not necessary, but do simplify our work a little.)

Now let $T$ be any tangent line in $S$. By (i) and (ii) there exists a sequence of lined $\{L_j\}$ converging to $T$ such that $n_{L_j}(f) < n_T(f)$ for all $j$. Suppose that $T$ is not tangent to the graph of $g$ anywhere. Then by (iii) we must have $n_{L_j}(g) = n_T(g)$ for large enough $j$. But this contradicts the fact that $n_{L_j}(f) = n_{L_j}(g)$ for all $j$, and that $n_T(f) = n_T(g)$, and hence every tangent line in $S$ must be tangent to the graph of $g$ somewhere. Reversing the roles of $f$ and $g$ we see that except for finitely many lines, $T(f)$ and $T(g)$ are the same. Since $f'$ and $g'$ are continuous, $T(f) = T(g)$, and the theorem follows by Theorem 2.

*Remark.* An interesting book in integral geometry is the one by Santalo **[3]**, where he defines the following metric:

$$d(f, g) = \frac{1}{2}\int |n_L(f) - n_L(g)|\, dL,$$

where $dL$ is a certain measure defined on sets of lines in the plane, and the integral is taken over all lines in the plane. The main reason we mention this here is that it is noted in **[3]** that $d(f, g) = 0$ implies that $f = g$ for a fairly general class of functions (actually more generally curves) $f$ and $g$. However no proof or reference is given for this result, and we simply note that this follows immediately from Theorem 3 with the assumptions given there on $f$ and $g$.

There is also a classical paper of Steinhaus **[4]** where he shows that if $n_L(f) = n_L(g)$ for every line $L$, then $f$ and $g$ must have the *same arc length*. Of course this

is not at all the same as saying that $f$ and $g$ are identical, though Steinhaus proves his result for a larger class of functions than we consider in the cutting line problem.

**5. Envelopes of families of lines.** In this section we discuss the connection between reconstructing $f$ from $T(f)$ and envelopes of families of lines. Given a one-parameter family of straight lines $F(x, y, t) = 0$, where $t$ is the parameter, an envelope is a function $f$ (or more generally, a curve) which has that family of lines as its tangent lines. When given the set $T(f)$ as a smooth *one-parameter family*, the method of reconstructing $f$ is well known and can be found in many advanced calculus or differential geometry texts (see, for example [**5**, §6.1]). However our approach is different because we consider $T(f)$ as a totally *unstructured* set of lines, without a specific parameterization!

Now it is customary in many texts to speak about a function being *the* envelope of its tangent lines, the implicit assumption being that there can only be *one* envelope. For example, in [**5**, §6.2], the author proves that a $C^2$ function $f$ is the envelope of its tangent lines, but there is nothing in his proof that prevents another function $g$ from also being an envelope of the *same* family of tangent lines! That is, two different parameterizations of the same family of lines could conceivably lead to two different envelopes. Put another way, must two functions with the same set of tangent lines be identical? But this is precisely the uniqueness problem we solved in Theorem 2, at least with the assumption that $f''$ and $g''$ have only finitely many zeros. We are able to prove Theorem 2 because our method of reconstructing $f$ does not depend on any particular parameterization of $T(f)$. Thus under appropriate smoothness assumptions it is permissible to refer to $f$ as *the* envelope of its set of tangent lines.

**6. Higher dimensions.** There are obvious ways to generalize our results to functions of more than one variable. The easy case is reconstructing, say, a space curve or a two-dimensional surface from its set of tangent *lines*. This can be reduced to the one-dimensional problem with little difficulty, and we leave the details to the reader. However the problem of reconstructing a surface $z = f(x, y)$ from its set of tangent *planes* is probably not reducible to the one variable case, and we leave this question for future research.

**7. Analytic functions on a domain.** In this section we show how to reconstruct an analytic function $f$ from its set of tangent lines relative to a domain $D$ in the complex plane. While $f$ is a function of two real variables, it is a function of just one complex variable, and thus the one-variable techniques of Section 3 will work here as well. The proofs are a little different, however, because there is no mean value theorem—though there is a power series expansion at our disposal. We define $T(f)$ in the obvious way as the set of linear functions $\{w = (f(z_0) - z_0 f'(z_0)) + zf'(z_0), z_0 \in D\}$. Assume that $f$ is not a linear function (otherwise the problem is trivial). We let $i(T_j, T) = (i_z(T_j, T), i_w(T_j, T))$ denote the intersection of $T_j$ and $T$ as a point in $C^2$. Here is the analog of Lemma 1 for analytic functions.

LEMMA 2. *Suppose $f$ is analytic in the closure of $D$, with $T_j$ and $T$ tangent lines at $(z_j, f(z_j))$ and $(z_0, f(z_0))$ respectively (assume all the $T_j$ are distinct with $m(T_j) \neq m(T)$ for all $j$). Suppose $z_j \to z_0$. Then $i(T_j, T) \to (z_0, f(z_0))$.*

*Proof.* As in the proof of Lemma 1 it suffices to show that $B_j \to 0$ ($B_j$ from (4) with $z_j$ and $z_0$ replacing $x_j$ and $x_0$). First expand $f$ and $f'$ in a power series about

$z_0$ and let $z = z_j$. A few computations give:

$$B_j = (z_j - z_0)\frac{\frac{1}{k!}f^{(k)}(z_0)(z_j - z_0)^{k-2} + \cdots}{\frac{1}{(k-1)!}f^{(k)}(z_0)(z_j - z_0)^{k-2} + \cdots},$$

where $k$ is the smallest integer $\geq 2$ such that $f^{(k)}(z_0) \neq 0$. It then follows that $B_j \to 0$ and the lemma is proved.

Now it follows easily that if $f$ is not a linear function, then $T(f)$ must contain *infinitely* many *distinct* tangent lines. For each of those tangent lines $T$ we obtain a point $(z_0, f(z_0))$ where $T$ is tangent using Lemma 2 and the approach from Section 3. Since we can recover infinitely many distinct points $(z_0, f(z_0))$ in this fashion, $f$ is uniquely determined by the Identity Theorem from complex variables. Note that we do not need the analog of Proposition 1 here, and indeed it might be false for analytic functions on a domain.

**8. Negative results.** We now consider again real-valued functions on an interval $I$, and show that Lemma 1 is false in general if $f''$ is allowed to have infinitely many zeros in $I$. It is not hard to construct a $C^\infty$ function $f$ with the following properties:

(i) $f(0) = f'(0) = 0$.

(ii) $f(x_j) = f'(x_j) \neq 0$ for some sequence of distinct numbers $\{x_j\}$ converging to 0.

We leave the details of the construction of $f$ to the reader.

Now let $T_j$ be the tangent line to $f$ at $(x_j, f(x_j))$, with $T =$ the $x$-axis being the tangent line at $(0,0)$. It is possible to construct $f$ so that the $T_j$ are all distinct. It follows that $i_x(T_j, T) = x_j - \dfrac{f(x_j)}{f'(x_j)} = x_j - 1 \to -1 \neq 0$, and thus $i(T_j, T) \not\to (0,0)$.

*Open Problems.*

1. Does $T(f) = T(g)$ imply $f \equiv g$ for any differentiable functions $f$ and $g$ defined on the same interval $[a, b]$? What about the entire real line?
2. Must a function that is analytic in the closure of a domain $D$ have at most finitely many multiple tangent lines?
3. Find a method for reconstructing a surface from its set of tangent planes. Does some analog of Lemma 1 hold in this case?
4. Must two differentiable functions $f(x, y)$ and $g(x, y)$ with the same set of tangent planes be identical?
5. If every tangent line to $f$ is multiple, must $f$ then be a straight line?
6. If every line in the plane intersects two continuous curves the same number of times, must the two curves be identical?

REFERENCES

1. J. W. Bruce and P. J. Giblin, Curves and Singularities, Cambridge University Press, 1984.
2. A. Horwitz and L. Rubel, Two theorems on inverse interpolation, *Rocky Mountain J. of Math.*, No. 3, (Summer 1988) 645–653.
3. L. Santalo, Integral Geometry and Geometric Probability, Encyclopedia of Mathematics and its Applications, vol. 1, Addison-Wesley, 1976.
4. H. Steinhaus, Sur la portée pratique et théorique de quelques théorèmes sur la measure des ensembles de droites, 1er Congrès des mathématiciens des pays slaves, Comptes-Rendus Du, Warszawa, 1929.
5. D. Widder, Advanced Calculus, 2nd edition, Prentice Hall, 1961.

*Proof.* Because $f$ is continuous and has compact domain, its graph $S$ is compact. Because the graph $S$ of $g$ is compact, $g$ is continuous.

*Note.* The fact that a function has compact domain and is continuous on that domain if and only if its graph is compact follows from theorem 12 of chapter 2 of **[1]**, which states

> Let $f$ be defined on a compact set $E$. Then, $f$ is continuous on $E$ if and only if the graph of $f$ is a compact set

together with the fact that if the graph of $f$ is compact then the domain of $f$ is compact. This is easily proved: if $\{x_n\}$ is any sequence of points in the domain, then $\{x_n, f(x_n)\}$ is a sequence of points of the graph, and, because the graph is compact, has a subsequence convergent to a point of the graph. Therefore $\{x_n\}$ has a subsequence convergent to a point of the domain. It follows that the domain is compact.

REFERENCE

1. R. C. Buck, Advanced Calculus, second edition, 1965 p. 75.

# Normalized Symmetric Functions, Newton's Inequalities, and a New Set of Stronger Inequalities

SHMUEL ROSSET
*Department of Mathematics, Tel Aviv University, Tel Aviv, Israel*

**1. Normalized symmetric functions.** If $x_1, \dots, x_n$ are $n$ numbers ('quantities'), their elementary symmetric functions are

$$e_1(x_1, \dots, x_n) = x_1 + \cdots + x_n$$

$$e_2(x_1, \dots, x_n) = x_1x_2 + \cdots + x_{n-1}x_n = \sum_{i<j} x_i x_j$$

and so on until

$$e_n(x_1, \dots, x_n) = x_1 \cdots x_n.$$

One knows that if $P(x)$ is defined by $P(x) = (x - x_1) \cdots (x - x_n)$, then $P(x) = x^n - e_1x^{n-1} + \cdots + (-1)^n e_n$.

For our purpose, it would be useful to *normalize* the elementary symmetric functions by dividing each $e_j$ by the number of its summands. We denote the normalized function by $E$. So

$$E_j(x_1, \dots, x_n) = \binom{n}{j}^{-1} e_j(x_1, \dots, x_n).$$

In this notation

$$P(x) = \sum_0^n (-1)^j \binom{n}{j} E_j(x_1, \dots, x_n) x^{n-j}, \qquad (E_0 = 1).$$

*Proof.* Because $f$ is continuous and has compact domain, its graph $S$ is compact. Because the graph $S$ of $g$ is compact, $g$ is continuous.

*Note.* The fact that a function has compact domain and is continuous on that domain if and only if its graph is compact follows from theorem 12 of chapter 2 of [**1**], which states

> Let $f$ be defined on a compact set $E$. Then, $f$ is continuous on $E$ if and only if the graph of $f$ is a compact set

together with the fact that if the graph of $f$ is compact then the domain of $f$ is compact. This is easily proved: if $\{x_n\}$ is any sequence of points in the domain, then $\{x_n, f(x_n)\}$ is a sequence of points of the graph, and, because the graph is compact, has a subsequence convergent to a point of the graph. Therefore $\{x_n\}$ has a subsequence convergent to a point of the domain. It follows that the domain is compact.

REFERENCE

1. R. C. Buck, Advanced Calculus, second edition, 1965 p. 75.

# Normalized Symmetric Functions, Newton's Inequalities, and a New Set of Stronger Inequalities

SHMUEL ROSSET
*Department of Mathematics, Tel Aviv University, Tel Aviv, Israel*

**1. Normalized symmetric functions.** If $x_1, \dots, x_n$ are $n$ numbers ('quantities'), their elementary symmetric functions are

$$e_1(x_1, \dots, x_n) = x_1 + \cdots + x_n$$

$$e_2(x_1, \dots, x_n) = x_1 x_2 + \cdots + x_{n-1} x_n = \sum_{i<j} x_i x_j$$

and so on until

$$e_n(x_1, \dots, x_n) = x_1 \cdots x_n.$$

One knows that if $P(x)$ is defined by $P(x) = (x - x_1) \cdots (x - x_n)$, then $P(x) = x^n - e_1 x^{n-1} + \cdots + (-1)^n e_n$.

For our purpose, it would be useful to *normalize* the elementary symmetric functions by dividing each $e_j$ by the number of its summands. We denote the normalized function by $E$. So

$$E_j(x_1, \dots, x_n) = \binom{n}{j}^{-1} e_j(x_1, \dots, x_n).$$

In this notation

$$P(x) = \sum_0^n (-1)^j \binom{n}{j} E_j(x_1, \dots, x_n) x^{n-j}, \qquad (E_0 = 1).$$

Note that if $x_1 = \cdots = x_n$ then $E_j(x_1, \ldots, x_n) = x_1^j$. Our first point in this note is to draw attention to the fact that there is a simple relationship between the $E$ functions of $P(x)$ and of its *derivative*. Let

$$Q(x) = \frac{1}{n}P'(x),$$

the 'normalized derivative'. Let the roots of $Q(x)$ be $y_1, \ldots, y_{n-1}$, so that

$$\begin{aligned} Q(x) &= (x - y_1) \cdots (x - y_{n-1}) \\ &= x^{n-1} - \binom{n-1}{1}E_1(y_1, \ldots, y_{n-1})x^{n-2} + \cdots + (-1)^{n-1}E_{n-1}(y_1, \ldots, y_{n-1}). \end{aligned}$$

But on the other hand

$$\begin{aligned} Q(x) &= \frac{1}{n}P'(x) \\ &= \sum_0^n (-1)^j \frac{n-j}{n}\binom{n}{j}E_j(x_1, \ldots, x_n)x^{n-j-1}. \end{aligned}$$

Equating coefficients of equal powers of $x$ gives for $j = 1, \ldots, n-1$

$$\boxed{E_j(x_1, \ldots, x_n) = E_j(y_1, \ldots, y_{n-1})}\,.$$

This formula seems not to be widely known. It is given for the case $j = 1$ in [3].

**2. The Newton inequalities.** The Newton inequalities, which are indeed due to Newton, say that if $x_1, \ldots, x_n$ are *real numbers* then for $j = 1, 2, \ldots, n-1$ (taking $E_0 = 1$)

$$E_j^2(x_1, \ldots, x_n) \geqslant E_{j-1}(x_1, \ldots, x_n) \cdot E_{j+1}(x_1, \ldots, x_n), \tag{1}$$

and the inequality is *strict* unless $x_1 = \cdots = x_n$, or both sides vanish. We will give a proof of these inequalities which is an inductive, and, therefore, very efficient variant of the proof of Hardy, Littlewood, and Pólya (see section 4.2 of [2]). In the next section we will show that this way of looking at the inequalities can be used to 'generate' new and *stronger* (and rather complicated!) inequalities.

To begin let $n = 2$. Then there is one inequality to prove:

$$(x_1 + x_2)^2 \geqslant 4x_1x_2$$

with equality if and only if $x_1 = x_2$. This is, of course, known to everyone. Now, assume $n > 2$ and the Newton inequalities proved (including the part referring to when equality holds) for all $(n-1)$-tuples of real numbers. Then, according to the notation of the last section for $j = 1, 2, \ldots, n-1$

$$E_j(x_1, \ldots, x_n) = E_j(y_1, \ldots, y_{n-1}), \tag{2}$$

where $y_1, \ldots, y_{n-1}$ are the roots of $Q(x)$.

We observe that (2) holds even if some of the roots $y_1, \ldots, y_{n-1}$ are complex. However, by Rolle's theorem $y_1, \ldots, y_{n-1}$ are real, are all within the closed interval $[\min(x_i), \max(x_i)]$ and $y_1, \ldots, y_{n-1}$ are not all equal if $x_1, \ldots, x_n$ are not all equal. Thus the inductive assumption applies and Newton's inequalities hold for

$E_1(y_1, \ldots, y_{n-1}), \ldots, E_{n-1}(y_1, \ldots, y_{n-1})$. By (2) we see that this already proves (1) for $j = 1, \ldots, n-2$ including the fact that (1) holds with strict inequality if $x_1, \ldots, x_n$ are not all equal, and both sides do not vanish.

It remains to prove (1) for $j = n-1$, that is,

$$E_{n-1}^2(x_1, \ldots, x_n) \geqslant E_{n-2}(x_1, \ldots, x_n) \cdot E_n(x_1, \ldots, x_n), \tag{3}$$

with equality iff $x_1 = \cdots = x_n$ *or* if both sides of (3) are zero. Recall that $E_n(x_1, \ldots, x_n) = x_1 \cdots x_n$. If some $x_i = 0$ (3) is clear as the right-hand side is zero. For $E_{n-1}(x_1, \ldots, x_n)$ also to be zero we need that some $x_k = 0$ too, where $k \neq i$. Then we see that if $x_i = 0$ for some $i$ then (3) holds either as a strict inequality or as $0 = 0$. Now assume all $x_i$ nonzero. Let $x_i' = 1/x_i$. It is easy to see that

$$E_j(x_1, \ldots, x_n)/E_n(x_1, \ldots, x_n) = E_{n-j}(x_1', \ldots, x_n'). \tag{4}$$

To prove that (3) holds strictly if not all $x_i$ are equal we divide it by $E_n^2(x_1, \ldots, x_n)$. We see that (3) is equivalent to

$$E_1^2(x_1', \ldots, x_n') \geqslant E_2(x_1', \ldots, x_n'),$$

which is true since $n > 2$. It is even true strictly (unless it reduces to $0 = 0$) since the $x_i'$ are not all equal.

This completes the proof of Newton's inequalities.

We observe that Maclaurin, Newton's student, had already noticed that these inequalities imply a very strong form of the arithmetic-geometric inequality. Assuming $x_i$ positive, Maclaurin proved

$$E_1(x_1, \ldots, x_n) \geqslant (E_2(x_1, \ldots, x_n))^{1/2} \geqslant \cdots \geqslant (E_n(x_1, \ldots, x_n))^{1/n} \tag{5}$$

and equality holds (even once) iff all $x_i$ are equal.

*Exercise.* Prove that if $E_1, \ldots, E_n$ are positive real numbers satisfying (1) then they satisfy (5). Show also that the converse is false. This exercise shows that the inequalities (5) are 'formal' consequences of (1) but are strictly *weaker*.

**3. A new set of inequalities.** The inequalities of Newton are 'quadratic'; they express the fact that if a polynomial $P(x)$ with real coefficients has only real zeros then the same is true of certain $n-1$ quadratic polynomials derived from it. More precisely, if $x_1, \ldots, x_n$ are real numbers and

$$P(x) = \prod_{i=1}^{n} (x - x_i) = \sum_{0}^{n} (-1)^j \binom{n}{j} E_j(x_1, \ldots, x_n) x^{n-j}.$$

then Newton's inequalities are just the statement that each of the quadratic polynomials

$$E_j(x_1, \ldots, x_n)x^2 - 2E_{j+1}(x_1, \ldots, x_n)x + E_{j+2}(x_1, \ldots, x_n) \qquad (j = 0, 1, \ldots, n-2) \tag{6}$$

has real roots.

From this point of view it is natural to ask: do analogous 'cubic' inequalities exist? We will now construct such a set of inequalities. These new inequalities are certainly more complicated (which may account for their not being known) but are

also strictly stronger. Recall that if

$$x^3 + ax^2 + bx + c \tag{7}$$

is a monic cubic polynomial, its discriminant $D$ (which, by definition, equals $(x_1 - x_2)^2(x_1 - x_3)^2(x_2 - x_3)^2$ where $x_1, x_2, x_3$ are the roots) is

$$D = 18abc - 4a^3c + a^2b^2 - 4b^3 - 27c^2. \tag{8}$$

Furthermore, it is easily checked that if the coefficients $a, b, c$ are real, then the roots $x_1, x_2, x_3$ are real iff $D \geqslant 0$ (cf. [**1**, p. 122]). Of course $D = 0$ iff two roots are equal.

It will be convenient to make changes of parameters in (7) and (8); so we will write a general monic cubic as

$$x^3 - 3\alpha x^2 + 3\beta x - \gamma. \tag{9}$$

Then $D$, as the function of $\alpha, \beta, \gamma$, is

$$D = 27\left(6\alpha\beta\gamma - 4\alpha^3\gamma + 3\alpha^2\beta^2 - 4\beta^3 - \gamma^2\right). \tag{10}$$

**The 'cubic' inequalities.** If $n \geqslant 3$ and $x_1, \ldots, x_n$ are real numbers then for $j = 0, \ldots, n-3$

$$6E_jE_{j+1}E_{j+2}E_{j+3} - 4E_jE_{j+2}^3 - E_j^2E_{j+3}^2 - 4E_{j+1}^3E_{j+3} + 3E_{j+1}^2E_{j+2}^2 \geqslant 0. \tag{11}$$

Here each $E_k$ stands for $E_k(x_1, \ldots, x_n)$ i.e., $\binom{n}{k}^{-1}e_k(x_1, \ldots, x_n)$. To prove (11) we again make an induction. If $n = 3$ then (11) is simply $D \geqslant 0$ for the cubic monic polynomial $(x - x_1)(x - x_2)(x - x_3)$. Now assume $n > 3$ and the inequalities (11) proved for all $(n-1)$-tuples of real numbers. As in §2 let

$$P(x) = (x - x_1)\cdots(x - x_n) = \sum_0^n(-1)^j\binom{n}{j}E_j(x_1, \ldots, x_n)x^{n-j}$$

$$Q(x) = \frac{1}{n}P'(x) = (x - y_1)\cdots(x - y_{n-1})$$

$$= \sum_0^{n-1}(-1)^j\binom{n-1}{j}E_j(y_1, \ldots, y_{n-1})x^{n-1-j}.$$

Since $E_j(y_1, \ldots, y_{n-1}) = E_j(x_1, \ldots, x_n)$ for $j = 0, \ldots, n-1$ (11) follows for $j = 0, \ldots, n-4$. It remains to prove (11) for $j = n-3$, that is,

$$6E_{n-3}E_{n-2}E_{n-1}E_n - 4E_{n-3}E_{n-1}^3 - E_{n-3}^2E_n^2 - 4E_{n-2}^3E_n + 3E_{n-2}^2E_{n-1}^2 \geqslant 0. \tag{12}$$

If $E_n = 0$, that is, some $x_i = 0$, this reduces to

$$3E_{n-2}^2E_{n-1}^2 \geqslant 4E_{n-3}E_{n-1}^3. \tag{13}$$

If $E_{n-1} = 0$ this is clear. If not, let us assume that the $x$'s are so numbered that

$x_n = 0$. Then

$$E_j(x_1,\ldots,x_n) = \binom{n}{j}^{-1} e_j(x_1,\ldots,x_n) = \binom{n}{j}^{-1} e_j(x_1,\ldots,x_{n-1})$$
$$= \frac{n-j}{n} E_j(x_1,\ldots,x_{n-1}).$$

Thus (13) is equivalent to (after cancelling $E_{n-1}^2(x_1,\ldots,x_n)$)

$$3 \cdot \frac{4}{n^2} E_{n-2}^2(x_1,\ldots,x_{n-1}) \geqslant 4 \cdot \frac{3}{n^2} \cdot E_{n-3}(x_1,\ldots,x_{n-1}) E_{n-1}(x_1,\ldots,x_{n-1})$$

and this is one of Newton's inequalities.

It remains to prove (12) in the case $E_n(x_1,\ldots,x_n) \neq 0$. Let $x_i' = 1/x_i$ and

$$E_j' = E_j(x_1',\ldots,x_n') = E_{n-j}/E_n.$$

Dividing (12) throughout by $E_n^4$ it becomes

$$6E_1'E_2'E_3' - 4E_1'^3E_3' + 3E_1'^2E_2'^2 - 4E_2'^3 - E_3'^2 \geqslant 0.$$

This last inequality holds because it is the inequality (11) with $j = 0$, for the $n$-tuple $x_1',\ldots,x_n'$ and this is true since $n > 3$. This completes the proof.

**Comparison of Newton's and the cubic inequalities.** It is interesting to note that if $E_0 = 1, E_1,\ldots,E_n$ are real numbers satisfying the inequalities (11) for $j = 0,1,\ldots,n-3$, then these numbers also satisfy Newton's inequalities $E_j^2 \geqslant E_{j-1} \cdot E_{j+1} (j = 1,\ldots,n-1)$. However, the converse is not true!

To see this note that if $E_j$, $E_{j+1}E_{j+2}$, $E_{j+3}$ satisfy (11) this means that the discriminant of

$$E_jx^3 - 3E_{j+1}x^2 + 3E_{j+2}x - E_{j+3}$$

is $\geqslant 0$. This implies it has 3 *real* roots (if $E_j \neq 0$. If $E_j = 0$ again (11) implies the existence of real roots.) Thus the coefficients satisfy Newton's inequalities.

To see that the converse does not hold it suffices to show that there exist 3 real numbers $a, b, c$ such that $c > 0$ and the coefficients of

$$R(x) = (x-a)(x-b-ic)(x-b+ic) \qquad (i = \sqrt{-1})$$

satisfy Newton's inequalities. The equations are

$$(a+2b)^2 \geqslant 3(b^2+c^2+2ab)$$
$$(b^2+c^2+2ab)^2 \geqslant 3a(a+2b)(b^2+c^2).$$

These inequalities are satisfied if $a = 1, b = -1, c = 1/2$. The polynomial $R(x)$, after clearing denominators, is $4x^3 + 4x^2 - 3x - 5$.

REFERENCES

1. S. Borofsky, Elementary Theory of Equations, Macmillan, New York, 1950.
2. G. H. Hardy, J. E. Littlewood, and G. Pólya, Inequalities, Cambridge University Press, 1959.
3. N. Levinson and R. Redheffer, Complex Variables, Holden Day, 1970.

# Reflection Sequences

N. ALON*, I. KRASIKOV, and Y. PERES
*Department of Mathematics, Tel Aviv University, Ramat-Aviv, Tel Aviv, Israel*

Let $X = (x_0, x_1, \dots, x_{n-1})$ be a circular sequence of real numbers and suppose their sum $S = \sum_{i=0}^{n-1} x_i$ is strictly positive. If, for some $i$, $x_i < 0$, then there is a *legal reflection* $R_i$ for $X$ defined as the operation which transforms $X$ into the circular sequence $X' = (x'_0, x'_1, \dots, x'_{n-1})$ obtained from $X$ by replacing $x_i$ by $-x_i$ and by subtracting $|x_i|$ from the two neighbors of $x_i$ in $X$. That is, $x'_i = -x_i$, $x'_{i-1} = x_{i-1} + x_i$, $x'_{i+1} = x_{i+1} + x_i$ and $x'_j = x_j$ for all $j, 0 \leqslant j < n$, $j \neq i-1, i, i+1$, where all indices are reduced modulo $n$. Notice that $\sum_{i=0}^{n-1} x'_i = \sum_{i=0}^{n-1} x_i$.

A *reflection sequence* for $X$ is a finite or infinite sequence $X = X_0, X_1, \dots$ of circular sequences and a sequence $R^1, \dots$ of reflections, such that $R^i$ is a legal reflection for $X_{i-1}$ that transforms it into $X_i$, $1 \leqslant i$. If all the elements of $X_t$ are nonnegative, that is, if there is no legal reflection for $X_t$, then we say that $X_t$ is *stable*.

If, for example, $n = 5$ and $X = (1, -2, -3, 8, 5)$, then one can easily check that $X = X_0$, $X_1 = (-1, 2, -5, 8, 5)$, $X_2 = (-1, -3, 5, 3, 5)$, $X_3 = (-4, 3, 2, 3, 5)$, $X_4 = (4, -1, 2, 3, 1)$ and $X_5 = (3, 1, 1, 3, 1)$ with the corresponding reflections $R_1, R_2, R_1, R_0, R_1$ a reflection sequence for $X$ which terminates in the stable configuration $X_5$.

A celebrated problem in the 1986 International Mathematical Olympiad (cf. **[1]**) asserts that for $n = 5$ and for any circular sequence of integers $X = (x_0, x_1, x_2, x_3, x_4)$ (whose sum is positive) any reflection sequence for $X$ must be finite. Equivalently, if we start from $X$ and apply, repeatedly, legal reflections, we eventually reach a stable circular sequence.

This result holds for general $n$ as stated in the following Proposition.

PROPOSITION 1. *If $X = (x_0, \dots, x_{n-1})$ is a circular sequence of $n(\geqslant 2)$ integers whose sum $S = \sum_{i=0}^{n-1} x_i$ is positive, then any reflection sequence for $X$ is finite.*

To prove Proposition 1 we need some notation. A sequence of consecutive numbers modulo $n$ is called an *arc*. We denote by $\langle i, k \rangle$ the arc $\langle i, k \rangle = (i, i+1, \dots, k-1, k)$ where all numbers are reduced modulo $n$. The *complementary arc* of $\langle i, k \rangle$ is $\langle k+1, i-1 \rangle$. For a sequence $X = (x_0, \dots, x_{n-1})$ and an arc $\langle i, k \rangle$ we define the arc-sum of $\langle i, k \rangle$ (with respect to $X$) by $S_X(\langle i, k \rangle) = \sum_{j \in \langle i, k \rangle} x_j$. Clearly, the sum of the arc-sum of any arc with the arc-sum of its complementary arc is precisely $\sum_{j=0}^{n-1} x_j = S$.

Suppose, now, that $X' = (x'_0, \dots, x'_{n-1})$ is obtained from $X$ by a legal reflection $R_i$. Clearly the arc-sum of any arc whose intersection with $\langle i-1, i+1 \rangle$ is of cardinality 0 or 3 does not change. The arc-sums of the two complementary arcs $\langle i, i \rangle$ and $\langle i+1, i-1 \rangle$ change from $x_i$ and $S - x_i$ to $-x_i$ and $S + x_i$, respectively. The arc-sums of other arcs that intersect $\langle i-1, i+1 \rangle$ in 1 or 2 elements only interchange, that is, $S_{X'}(\langle k, i \rangle) = S_X(\langle k, i-1 \rangle)$, $S_{X'}(\langle i, k \rangle) = S_X(\langle i+1, k \rangle)$, $S_{X'}(\langle k, i-1 \rangle) = S_X(\langle k, i \rangle)$ and $S_{X'}(\langle i+1, k \rangle) = S_X(\langle i, k \rangle)$.

*Research supported in part by an Allon Fellowship and by a grant from Bat Sheva de-Rothschild Foundation.

Define $f(X) = \Sigma|S_X(\langle i, k\rangle)|^2$ where the summation ranges over all arcs $\langle i, k\rangle$. By the preceding paragraph

$$f(X') - f(X) = (S + x_i)^2 + x_i^2 - (S - x_i)^2 - x_i^2 = 4Sx_i < 0.$$

Therefore, if $X = X_0, X_1, X_2, \dots$ is a reflection sequence for $X$, then $(f(X_i))_{i \geqslant 1}$ is a strictly decreasing sequence of nonegative integers and thus must be finite. This completes the proof of Proposition 1. □

The conclusion of Proposition 1 clearly holds even if we replace the assumption that $X$ is a sequence of integers by the assumption that it is a sequence of rationals. (Simply write each $x_i$ as an integral multiple of some rational $q$ and apply the result for integers.) The proof, however, does not imply finiteness for general real sequences. We next show that finiteness, in fact, holds for every real circular sequence $X$. Moreover, any two reflection sequences for $X$ that terminate in stable configurations have the same length.

PROPOSITION 2. *Let $X = (x_0, \dots, x_{n-1})$ be a circular sequence of $n(\geqslant 2)$ real numbers whose sum $S = \Sigma_{i=0}^{n-1} x_i$ is positive. Then any reflection sequence for $X$ is finite. Moreover, the length of any reflection sequence for $X$ that terminates in a stable configuration is independent of the sequence (and can be easily computed from the $x_i$'s).*

*Proof.* Associate each pair of complementary arcs $\lambda = \{\langle i, k\rangle, \langle k+1, i-1\rangle\}$ with a pair of real numbers $\{a_\lambda, b_\lambda\}$, where $a_\lambda = S_X(\langle i, k\rangle)$ and $b_\lambda = S_X(\langle k+1, i-1\rangle)$. Clearly $a_\lambda + b_\lambda = S$. If, for some $\lambda$, either $a_\lambda$ or $b_\lambda$ is negative, say, $a_\lambda < 0$, we define a *legal switch* $Q_\lambda$ as the operation that transforms $\{a_\lambda, b_\lambda\}$ into $\{-a_\lambda, b_\lambda + 2a_\lambda\}$ and does not change all the other pairs. A set of pairs $\{\{a_\lambda, b_\lambda\}\}$ is *stable* if $a_\lambda, b_\lambda \geqslant 0$ for all $\lambda$, that is, if there is no legal switch for it. Obviously the operators $Q_\lambda$ commute. The key idea is that any reflection sequence corresponds to a sequence of legal switches on the pairs $\{a_\lambda, b_\lambda\}$. Indeed, by the arguments given in the proof of Proposition 1, any legal reflection $R_i$ corresponds to a legal switch on the pair $\lambda = \{\langle i, i\rangle, \langle i+1, i-1\rangle\}$, together with a permutation on the other pairs. Therefore, the only effect of the legal reflection $R_i$ on the multiset of pairs $\{a_\lambda, b_\lambda\}$, where $\lambda$ ranges over all pairs of complementary arcs, is obtained by applying the legal switch $Q_\lambda$ for $\lambda = \{\langle i, i\rangle, \langle i+1, i-1\rangle\}$.

Returning to the commuting operators $Q_\lambda$ it is easy to check that if $a_\lambda < 0$ and one can apply $k$ repeated legal switches $Q_\lambda$ to $\{a_\lambda, b_\lambda\}$, then

$$Q_\lambda^k\{(a_\lambda, b_\lambda)\} = \begin{cases} \{kb_\lambda + (k+1)a_\lambda, -(k-1)b_\lambda - ka_\lambda\} & k \text{ even} \\ \{-ka_\lambda - (k-1)b_\lambda, kb_\lambda + (k+1)a_\lambda\} & k \text{ odd.} \end{cases}$$

It follows that the number of consecutive legal switches that can be applied to a pair $\{a_\lambda, b_\lambda\}$ with $a_\lambda < 0$ is precisely

$$\min\{k \geqslant 1 : kb_\lambda + (k+1)a_\lambda \geqslant 0\} = \lceil |a_\lambda|/(b_\lambda + a_\lambda)\rceil = \lceil |a_\lambda|/S\rceil.$$

As the operators $Q_\lambda$ commute this implies that the length of any sequence of legal switches that terminates in a stable configuration is precisely

$$\sum_{a_\lambda < 0} \left\lceil \frac{|a_\lambda|}{S} \right\rceil + \sum_{b_\lambda < 0} \left\lceil \frac{|b_\lambda|}{S} \right\rceil. \tag{1}$$

REFERENCES

1. *Mathematics Magazine*, 59 (1986), 253–254.
2. S. Mozes, Reflection processes on graphs and Weyl groups, preprint (1987).

# Counting the Rationals

YORAM SAGHER
*Department of Mathematics, Statistics, and Computer Science, University of Illinois, Chicago, IL*

Cantor's proof of the countability of the positive rationals has great appeal. One sees the idea literally at a glance. On the other hand the construction counts all ordered pairs of positive integers so that each positive rational is counted infinitely many times, and if one wants, say, the $10^{15}$th positive rational in Cantor's list, one has to keep counting for a considerable time.

Here we offer a direct way of counting the positive rationals. Given $m/n$, we can assume that $m$ and $n$ are relatively prime. Let $m = p_1^{e_1} \cdots p_k^{e_k}$, $n = q_1^{f_1} \cdots q_l^{f_l}$, be the prime-number decompositions of $m$ and $n$. The counting function is defined by: $f(1) = 1$ and

$$f\left(\frac{m}{n}\right) = p_1^{2e_1} \cdots p_k^{2e_k} q_1^{2f_1-1} \cdots q_l^{2f_l-1}.$$

$f$ is clearly 1-1 and onto. The $10^{15}$th positive rational in this list is $10^{-8}$.

REFERENCES

1. *Mathematics Magazine*, 59 (1986), 253–254.
2. S. Mozes, Reflection processes on graphs and Weyl groups, preprint (1987).

# Counting the Rationals

YORAM SAGHER
*Department of Mathematics, Statistics, and Computer Science, University of Illinois, Chicago, IL*

Cantor's proof of the countability of the positive rationals has great appeal. One sees the idea literally at a glance. On the other hand the construction counts all ordered pairs of positive integers so that each positive rational is counted infinitely many times, and if one wants, say, the $10^{15}$th positive rational in Cantor's list, one has to keep counting for a considerable time.

Here we offer a direct way of counting the positive rationals. Given $m/n$, we can assume that $m$ and $n$ are relatively prime. Let $m = p_1^{e_1} \cdots p_k^{e_k}$, $n = q_1^{f_1} \cdots q_l^{f_l}$, be the prime-number decompositions of $m$ and $n$. The counting function is defined by: $f(1) = 1$ and

$$f\left(\frac{m}{n}\right) = p_1^{2e_1} \cdots p_k^{2e_k} q_1^{2f_1-1} \cdots q_l^{2f_l-1}.$$

$f$ is clearly 1-1 and onto. The $10^{15}$th positive rational in this list is $10^{-8}$.

THE TEACHING OF MATHEMATICS

EDITED BY MELVIN HENRIKSEN AND STAN WAGON

# The Radius of Convergence of Power Series Solutions to Linear Differential Equations

ISOM H. HERRON
*Department of Mathematics, Howard University, Washington, DC 20059*

When the coefficients of a linear ordinary differential equation are not all constants, there exist few methods for determining the solution in general. One method that still has great currency is to obtain the solution by power series. Though most of the ideas of this note apply to equations of arbitrary order, we will concentrate on the second-order equation typified by

$$y'' + p(x)y' + q(x)y = 0. \tag{1}$$

The important fact about (1) is that the behavior of its solutions near a point $x_0$ is determined by the behavior of $p(x)$ and $q(x)$ near this point. If $p(x)$ and $q(x)$ are *analytic* at $x_0$, that is, each has a power series convergent in some neighborhood of $x = x_0$, then $x = x_0$ is an *ordinary point* of the equation. A solution is sought of the form

$$y = \sum_{n=0}^{\infty} a_n(x - x_0)^n. \tag{2}$$

For instance, if we consider equation (1) with $p(x) = 3x/(x^2 + 4)$ and $q(x) = 1/(x^2 + 4)$, or in working form

$$(x^2 + 4)y'' + 3xy' + y = 0, \tag{3}$$

the substitution $y = \sum_{n=0}^{\infty} a_n x^n$ leads to

$$a_{n+2} = -\frac{(n+1)}{4(n+2)}a_n.$$

There is no problem in finding the first few terms in the series, but treating convergence is more delicate. Most textbooks rely on something like the following:

THEOREM A [2, p. 155]. *Let $x_0$ be an ordinary point of* (1), *and let $a_0$ and $a_1$ be arbitrary constants. Then there exists a unique function of the form* (2) *that is analytic at $x = x_0$, is a solution of* (1) *in a certain neighborhood of this point, and satisfies the initial conditions $y(x_0) = a_0$, $y'(x_0) = a_1$. Furthermore, if the power series expansions of $p(x)$ and $q(x)$ are valid on an interval $|x - x_0| < R$, $R > 0$, then the power series expansion* (2) *is also valid on this interval.*

The proof of this theorem is also given in the cited reference. However, most introductory courses in differential equations do not prove this theorem and so the students must, as often is the case, apply its conclusions on faith.

In this note, we point out how the radius of convergence of the series (2) may be found from the recurrence formula without recourse to Theorem A, in the case when

the formula involves two terms such as

$$a_{n+p} = f(n)a_n, \qquad n = 0, 1, 2, \dots . \tag{4}$$

These are grouped naturally, for instance as odd or as even terms when $p = 2$. In the general case we make use of the following:

THEOREM B. *Suppose the series* (2) *is generated by the recurrence formula* (4), *then the series converges at least for* $|x - x_0| < R$, *where*

$$R = \left[\lim_{k\to\infty} |f(pk)|\right]^{-1/p}. \tag{5}$$

*Proof.* We notice from (2) and (4) that $a_n$ and $a_{n+p}$ are successive terms in $p$ "grouped" power series

$$\sum_{k=0}^{\infty} a_{pk+s}(x - x_0)^{pk+s}, \qquad s = 0, 1, 2, \dots, p - 1,$$

where $n = pk + s$. (As solutions of (1), where $p \geq 2$, only two of these series are nonvoid.) Begining with $s = 0$, one observes then that using the ratio test, the limiting absolute value of the ratio of successive terms is

$$\lim_{k\to\infty}\left[\left|\frac{a_{pk+p}}{a_{pk}}\right| |x - x_0|^p\right] = \lim_{k\to\infty}\left[|f(pk)|\,|x - x_0|^p\right] = f_p|x - x_0|^p,$$

where

$$f_p = \lim_{k\to\infty} |f(kp)|,$$

assuming that limit exists, so the series converges when

$$f_p|x - x_0|^p < 1$$

or

$$|x - x_0| < \frac{1}{f_p^{1/p}}.$$

Likewise, when $s = 1$, the absolute value of the ratio of successive terms of (2) is, in the limit,

$$\lim_{k\to\infty}\left[\left|\frac{a_{pk+p+1}}{a_{pk+1}}\right| |x - x_0|^p\right] = \lim_{k\to\infty}\left[|f(pk + 1)|\ |x - x_0|^p\right].$$

Since

$$\lim_{k\to\infty} |f(pk + 1)| = \lim_{k\to\infty} |f(pk)|,$$

result (5) is again obtained. The limits of the other ratios for $s = 2, \dots, p - 1$ are

$$\lim_{k\to\infty} |f(pk + s)| = \lim_{k\to\infty} |f(pk)| = f_p, \qquad s = 2, \dots, p - 1$$

which again leads to (5).

The proof for the case $s = 0$ is contained in [**1**, p. 120].

REFERENCES

1. F. B. Hildebrand, Advanced Calculus for Applications, 2nd ed., Prentice-Hall, Englewood Cliffs, N.J., 1976.
2. G. F. Simmons, Differential Equations, McGraw-Hill, New York, 1972.

# Pursuing Analogies Between Differential Equations and Difference Equations

DAVID L. ABRAHAMSON
*Department of Mathematics and Computer Science, Rhode Island College, Providence, RI 02908*

**1. Introduction.** The study of ordinary differential equations has long been a staple in mathematics at both the undergraduate and graduate levels. Recently, instruction in the study of difference equations has widened, primarily due to the expanded role of the digital computer in mathematics. The two topics are inextricably linked at all levels, from elementary techniques through current research questions. Pursuing the analogies between these fields of study can only deepen the understanding of each. In particular, the study of many elementary topics in difference equations, requiring not even the use of calculus, can serve as a foundation for intuition and understanding of the analogous topics in differential equations. Since typical difficulties encountered at the introductory level in studying differential equations include development of intuition and avoiding the approach of pure memorization of formulas, such a foundation is indeed useful.

The purpose of this article is to illustrate how the analogy might be pursued through some typical problems and to comment on the usefulness of this analogy in the study of other topics. The comments here are, of course, not intended to be exhaustive; the goal is to suggest another way of thinking along with the traditional techniques for studying differential equations.

Note here that a thorough introduction to linear difference equations is provided by Miller [**5**], while Cadzow [**1**], Charlton [**2**], May [**4**], and others have described some of the dynamics and applications of first-order and higher-order difference equations.

**2. Initial-value problems.** Consider

$$y_{n+1} = F(n, y_n), \qquad n = 0, 1, 2, \ldots, \tag{1}$$

and

$$\dot{y}(t) = G(t, y(t)) \tag{2}$$

for $t$ in some interval containing 0. To the student encountering the ordinary differential equation (2) for the first time, the notion that it may have many solutions, and that we appeal to a so-called initial value $y(0) = y_0$ to distinguish among solutions, is often foreign. On the other hand, the difference equation (1) can very easily be shown, by example, to exhibit such properties. A simple equation such as $y_{n+1} = y_n^2$ or $y_{n+1} = \cos y_n$ can quickly be investigated using a calculator. It will be apparent that many solution sequences exist (and can have widely varying

REFERENCES

1. F. B. Hildebrand, Advanced Calculus for Applications, 2nd ed., Prentice-Hall, Englewood Cliffs, N.J., 1976.
2. G. F. Simmons, Differential Equations, McGraw-Hill, New York, 1972.

# Pursuing Analogies Between Differential Equations and Difference Equations

DAVID L. ABRAHAMSON
*Department of Mathematics and Computer Science, Rhode Island College, Providence, RI 02908*

**1. Introduction.** The study of ordinary differential equations has long been a staple in mathematics at both the undergraduate and graduate levels. Recently, instruction in the study of difference equations has widened, primarily due to the expanded role of the digital computer in mathematics. The two topics are inextricably linked at all levels, from elementary techniques through current research questions. Pursuing the analogies between these fields of study can only deepen the understanding of each. In particular, the study of many elementary topics in difference equations, requiring not even the use of calculus, can serve as a foundation for intuition and understanding of the analogous topics in differential equations. Since typical difficulties encountered at the introductory level in studying differential equations include development of intuition and avoiding the approach of pure memorization of formulas, such a foundation is indeed useful.

The purpose of this article is to illustrate how the analogy might be pursued through some typical problems and to comment on the usefulness of this analogy in the study of other topics. The comments here are, of course, not intended to be exhaustive; the goal is to suggest another way of thinking along with the traditional techniques for studying differential equations.

Note here that a thorough introduction to linear difference equations is provided by Miller [**5**], while Cadzow [**1**], Charlton [**2**], May [**4**], and others have described some of the dynamics and applications of first-order and higher-order difference equations.

**2. Initial-value problems.** Consider

$$y_{n+1} = F(n, y_n), \qquad n = 0, 1, 2, \dots, \tag{1}$$

and

$$\dot{y}(t) = G(t, y(t)) \tag{2}$$

for $t$ in some interval containing 0. To the student encountering the ordinary differential equation (2) for the first time, the notion that it may have many solutions, and that we appeal to a so-called initial value $y(0) = y_0$ to distinguish among solutions, is often foreign. On the other hand, the difference equation (1) can very easily be shown, by example, to exhibit such properties. A simple equation such as $y_{n+1} = y_n^2$ or $y_{n+1} = \cos y_n$ can quickly be investigated using a calculator. It will be apparent that many solution sequences exist (and can have widely varying

behaviors); the initial value $y_0$ in (1) is needed to specify a particular solution. Thus the existence and uniqueness of the solution to the initial-value problem (IVP) consisting of (1) along with a given value of $y_0$ is completely obvious. A student will, immediately and intuitively, understand "starting at $y_0$" and moving from there as determined by the equation. By appeal to this line of thinking, the basic idea of a standard existence-uniqueness result for the IVP consisting of (2) with a given value $y(0) = y_0$ is plausible even to the complete novice; one "starts at $y_0$" and moves from there as determined by (2).

In an advanced course, the question of the maximal interval of existence of the solution to an IVP for (2) will be addressed. To establish intuition, we may again appeal to a simple example in (1); for instance, the IVP

$$y_{n+1} = \frac{2}{(1 - y_n)}, \qquad y_0 = 3,$$

generates terms $y_1 = -1$, $y_2 = 1$, and its solution then fails to exist further because the value $y_2 = 1$ is not in the domain of the right-hand side of the equation. Since the standard results concerning the interval of existence of the solution to an IVP for (2) essentially center on whether the solution approaches the boundary of the domain of the right-hand side of (2), the analogy is helpful.

At any level of study, autonomous versions of (1) and (2), namely,

$$y_{n+1} = f(y_n) \tag{3}$$

and

$$\dot{y}(t) = g(y(t)) \tag{4}$$

are of special importance. Note here that one may again appeal to simple examples in (1) and (3) to illustrate the contrast between autonomous and nonautonomous equations and hence to carry that idea to (2) and (4).

**3. Linear equations.** In an introductory study of differential equations, the solution of first-order linear equations is a standard early topic. Because students are unfamiliar with the methods involved and sometimes have not retained all of the subtleties from calculus, they often merely memorize some of the formulas that are presented. The corresponding study in difference equations is simpler and more easily remembered and therefore provides a useful foundation.

We might consider

$$y_{n+1} = ay_n + p_n \tag{5}$$

and

$$\dot{y}(t) = by(t) + q(t), \tag{6}$$

where $a$ and $b$ are constants, $\{p_n\}$ is a given sequence, and $q$ is a given function. The solution to an IVP for (5) can be obtained by direct calculation and guessing: from $y_1 = ay_0 + p_0$, $y_2 = a^2y_0 + ap_0 + p_1$, etc., we predict that

$$y_n = a^n y_0 + \sum_{k=0}^{n-1} a^{n-1-k} p_k \tag{7}$$

and then verify that this is the correct formula. Whatever method and terminology is used to solve the IVP consisting of (6) along with a given value $y(0) = y_0$, the formula

$$y(t) = e^{bt}y_0 + \int_0^t e^{b(t-s)}q(s)\,ds \tag{8}$$

is obtained. The similarity between (7) and (8) is striking and useful. For example, the numbers $a$ and $e^b$ share a common role and the summation and convolution-type integral bear a powerful resemblance to one another. The details of the correspondence may be pursued at length; one may address homogeneous equations, stability, the role of the initial value, and so on. The importance of pointing out the correspondences lies in the fact that (5) and (7) are so simple to understand.

**4. Linear vs. nonlinear.** At an introductory level, even very able students can be slow to appreciate the depth of the contrast between problems involving linear differential equations and those involving nonlinear ones. A few examples of nonlinear difference equations can illustrate the fundamental points that need to be appreciated. To list only a few here, we note that closed-form solutions may not exist or may be elusive, growth or decay may not be of exponential nature, equilibria may have various properties of attraction or repulsion, and so on. Examples such as $y_{n+1} = k_1 y_n(k_2 - y_n)$, $k_1$, $k_2$ constant, suggest these ideas and more.

**5. Higher-order equations and advanced topics.** As with (1) and (2), we may compare

$$y_{n+2} = H(n, y_n, y_{n+1}) \tag{9}$$

with

$$\ddot{y}(t) = K(t, y(t), \dot{y}(t)) \tag{10}$$

and we can study pairs of higher-order equations. We motivate the appropriate IVP for (10) by appealing to specific examples of (9). The difficulty of another problem involving (10)—say, a two-point boundary value problem—can also be illustrated by using (9).

The standard approach to studying (10) begins with linear equations and includes work on two-dimensional first-order systems. Both have appropriate analogues among difference equations. We might compare

$$ay_{n+2} + by_{n+1} + cy_n = p_n \tag{11}$$

with

$$\alpha\ddot{y}(t) + \beta\dot{y}(t) + \gamma y(t) = q(t) \tag{12}$$

where $a$, $b$, $c$, $\alpha$, $\beta$, and $\gamma$ are constants, $\{p_n\}$ is a given sequence, and $q$ is a given function. The pursuit of solutions to the homogeneous versions of (11) and (12)—of the form $\{r^n\}$ for (11) and $e^{st}$ for (12), based on first-order experience—leads to characteristic equations for each. Conclusions about behavior of solutions to the homogeneous equations can be made based on the nature of the roots of the characteristic polynomials. The importance of linearity and the appropriateness of

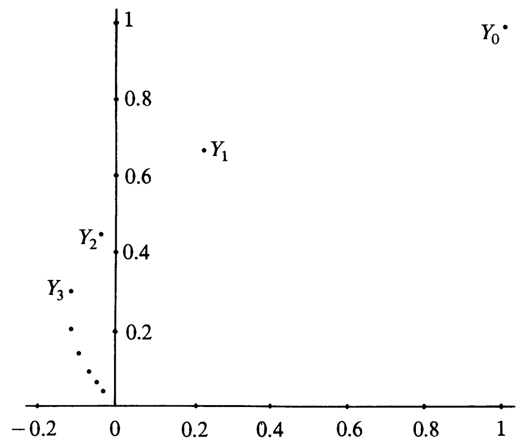



FIG. 1

The array of topics to be addressed in a more advanced course is wide and is a matter of the instructor's preference. Although more experienced students are in less need of support for their intuition, the parallel study of difference equations along with differential equations remains of pedagogical use and is of interest in its own right. Many topics in differential equations have direct analogues in difference equations, and, as has been suggested here, the corresponding study may well be simple and rewarding.


REFERENCES

1. J. A. Cadzow, Discrete-Time Systems, An Introduction with Interdisciplinary Applications, Prentice-Hall, Englewood Cliffs, N.J., 1973.
2. F. Charlton, Ordinary Differential and Difference Equations, Theory and Applications, Van Nostrand, London, 1965.
3. J. Hale, Ordinary Differential Equations, Wiley, New York, 1969.
4. R. M. May, Simple mathematical models with very complicated dynamics, *Nature*, 261 (1976) 459–467.
5. K. S. Miller, Linear Difference Equations, W. A. Benjamin, New York, 1968.


# On the Use of Iteration Methods for Approximating the Natural Logarithm


JAMES F. EPPERSON
*Department of Mathematics and Statistics, University of Alabama, Huntsville, AL 35899*


**1. Introduction.** Usually one generates approximations to functions such as $\log(x)$ and $\exp(x)$ by first doing some kind of range reduction and then employing a polynomial or rational approximation. Reference [2] gives a number of examples of this, some of which have been actually used on some computers [3]. In this note

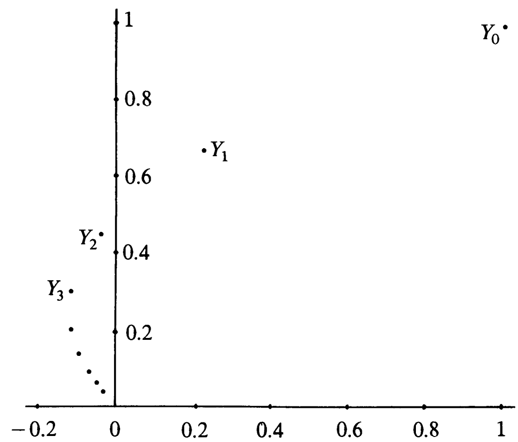



FIG. 1

The array of topics to be addressed in a more advanced course is wide and is a matter of the instructor's preference. Although more experienced students are in less need of support for their intuition, the parallel study of difference equations along with differential equations remains of pedagogical use and is of interest in its own right. Many topics in differential equations have direct analogues in difference equations, and, as has been suggested here, the corresponding study may well be simple and rewarding.

REFERENCES

1. J. A. Cadzow, Discrete-Time Systems, An Introduction with Interdisciplinary Applications, Prentice-Hall, Englewood Cliffs, N.J., 1973.
2. F. Charlton, Ordinary Differential and Difference Equations, Theory and Applications, Van Nostrand, London, 1965.
3. J. Hale, Ordinary Differential Equations, Wiley, New York, 1969.
4. R. M. May, Simple mathematical models with very complicated dynamics, *Nature*, 261 (1976) 459–467.
5. K. S. Miller, Linear Difference Equations, W. A. Benjamin, New York, 1968.

# On the Use of Iteration Methods for Approximating the Natural Logarithm

JAMES F. EPPERSON
*Department of Mathematics and Statistics, University of Alabama, Huntsville, AL 35899*

**1. Introduction.** Usually one generates approximations to functions such as $\log(x)$ and $\exp(x)$ by first doing some kind of range reduction and then employing a polynomial or rational approximation. Reference [2] gives a number of examples of this, some of which have been actually used on some computers [3]. In this note

we wish to discuss—at a level appropriate for a junior/senior course in numerical analysis—the computation of some elementary functions by *iteration*. While some of the examples appear to be competitive in terms of computational efficiency, our main purpose is to present an interesting application that brings together several aspects of an introductory numerical analysis course.

The basic idea is the following: let $g(x)$ and $G(x)$ be a function/inverse pair, i.e., $G(g(x)) = g(G(x)) = x$ for all $x$ in appropriate domains. If one of these—say, $g(x)$—can be easily computed (by any means) then the other can be computed by finding the root of the equation

$$g(\alpha) - x = 0. \tag{1}$$

In other words, if $\alpha$ solves (1), then $G(x) = \alpha$. We have thus converted a problem in approximation of functions to a problem in root-finding. The canonical example is the square root of $x$, which can be computed via the iteration (derived from Newton's Method):

$$\alpha_{n+1} = \frac{1}{2}\left(\alpha_n + \frac{x}{\alpha_n}\right).$$

**2. Logarithm.** Among the elementary transcendental functions, the logarithm is in many ways one of the hardest to approximate, because its Taylor series—a logical starting point for many approximations—is so slow to converge. Even though one can reduce the problem of computing $\log(x)$ for $x$ on $(0,\infty)$ to that of computing $\log(\xi)$ for $\xi \in [\frac{1}{2}, 1]$, it still requires a fair amount of computation to get $\log(\xi)$.

Following the idea outlined above, we wish to solve the equation

$$f(\alpha) = e^{\alpha} - x = 0.$$

That is, if $\alpha$ satisfies the above, then $\alpha = \log(x)$. Applying Newton's Method yields the iteration

$$\alpha_{n+1} = \alpha_n + x\exp(-\alpha_n) - 1, \tag{2}$$

which can be carried out without having to compute $\log(x)$. At this point it is appropriate to ask: For what values of $\alpha_0$ does this iteration converge, and how fast? To rephrase the question, can we find a good enough initial approximation to $\alpha = \log x$ so that (2) converges rapidly?

The usual error expression for Newton's Method is (see [**1**]):

$$e_{n+1} = \frac{-e_n^2 f''(\xi_n)}{2\, f'(\alpha_n)}$$

Here $e_n = \alpha - \alpha_n$ and $\xi_n$ is between $\alpha$ and $\alpha_n$. If we let $\varepsilon_n = |e_n|$ then we can produce the error bound:

$$\varepsilon_{n+1} \leq \frac{\varepsilon_n^2 \exp(\varepsilon_n)}{2}. \tag{3}$$

If we define "convergence after $n$ iterations" to mean that $\varepsilon_n \leq 2^{-48}$, then we get convergence in about 4 iterations from an initial error of about 0.15. It remains to get an initial guess of known accuracy.

To begin with we note that we can assume, with no loss of generality, that $x$ is in the form $x = 2^m\xi$ for $m$ an integer and $\xi \in [\frac{1}{2}, 1]$. Thus

$$\log(x) = m\log(2) + \log(\xi),$$

and we need only compute the logarithm for a relatively small set of values.

As a first attempt at generating an initial guess, consider linear interpolation to the logarithm over the interval $[1/2, 1]$. If $\ell(x)$ is the linear interpolate, then we quickly have that

$$\ell(x) = 2(x-1)\log(2)$$

and

$$|\log(x) - \ell(x)| \le \frac{1}{8}$$

the latter result following quickly from the usual estimate for linear interpolation [**1**]. Therefore, if we take $\alpha_0 = \ell(x)$, we are guaranteed that the initial error is less than 0.125. From (3), then, we should expect to get convergence in about four iterations. Note that we can rearrange $\ell(x)$ so that it can be computed in only 2 floating point operations, or *flops*.

A second alternative would be Hermite interpolation over $[1/2, 1]$. This yields (see [**1**] for the Hermite error estimate):

$$h(x) = (x-1)\left(A + (x-1)\left(B + C\left(x - \frac{1}{2}\right)\right)\right),$$

where $A = 1$, $B = 2 - 4\log(2)$, and $C = 12 - 16\log(2)$; then

$$|\log(x) - h(x)| \le \frac{1}{64} = 0.015625.$$

Convergence now comes in about 3 iterations. This is probably the better of the two initial estimates, since the savings of a single iteration more than offsets the slightly higher cost of actually computing the initial guess. (The cost of computing $h(x)$ is 7 flops.) In fact, we can quickly calculate that the cost of computing $\log(x)$ using the Hermite initial guess is 16 flops plus $3\exp(\cdot)$ calls. If the simple linear interpolate were used, the cost would be 14 flops plus $4\exp(\cdot)$ calls.

Iteration does not work very well going the other way, however. If we assume that $\log(x)$ is easy to compute, and try to use Newton's Method to compute $\exp(x)$, then we must find the root of the function

$$f(\alpha) = \log(\alpha) + x,$$

so that $\alpha = \exp(-x)$. Newton's Method applied to this function yields the iteration

$$\alpha_{n+1} = \alpha_n(1 - x - \log(\alpha_n)) \tag{4}$$

and the error expression yields the bound $\varepsilon_{n+1}|\varepsilon_n^2/2|\alpha_n|$, which is somewhat more pessimistic than (3), because the coefficient of quadratic convergence may be quite large. In fact, the iteration (4) may not even be defined, since it is possible for $\alpha_n$ to become negative. Convergence now requires much more accurate initial values, and the practicality of the scheme is doubtful.

particularly appropriate to point out that one advantage of approximation via iteration is that the accuracy of the approximation is limited only by the arithmetic of the machine doing the work. This is in marked contrast to polynomial or rational approximations, for example, which must usually be completely reconstructed in order to improve accuracy.

REFERENCES

1. Kendall E. Atkinson, An Introduction to Numerical Analysis, Wiley, New York, 1978.
2. J. Hart, et al., Computer Approximations, Wiley, New York, 1968.
3. Fortran Version 5 Common Library Mathematical Routines Reference Manual, Control Data Corporation, 1979.
4. Walter Gander, On Halley's iteration method, this MONTHLY (92) 131–134.
5. George H. Brown, Jr., On Halley's variation of Newton's method, this MONTHLY (84) 726–728.
6. J. M. Borwein and P. B. Borwein, The arithmetic-geometric mean and fast computation of elementary functions, *SIAM Rev.*, 26, 351–366.
7. ______, Pi and the AGM, Wiley-Interscience, New York, 1987.

# There Are No Safe Virus Tests

WILLIAM F. DOWLING
*Department of Mathematics and Computer Science, Drexel University, Philadelphia, PA 19104*

This note gives a proof that no program can both test its input for the presence of a virus and simultaneously be guaranteed not to spread a virus itself. (You may define "virus" any way you please, as long as the definition is extensional.) This immediate corollary of Rice's Theorem [**1**] is proved by a direct diagonalization and offered as an antidote (not a vaccine) to boredom in the elementary computability course during the presentation of the halting problem.

Programs running on modern computers, unlike the executions of Turing machine programs, as usually conceived, run "in an environment." This is to say that when a program is executed, it is run as a subprogram of the logically independent program, the operating system, which is responsible for such bookkeeping chores as primary and secondary memory management, process management, recording statistics, and so on. A program that, when run, alters the code of the operating system, is called a *virus*. (When a new version of the operating system is written legitimately, it replaces, not alters, the former operating system.) This is a somewhat less restrictive definition of virus than others have proposed [**2**], in that no particular behavior is required of the modified operating system. For instance, it is frequently required that a virus have the effect of inserting its own code into other executable programs. Such restrictions are unnecessary for the result we seek.

It would be nice if we could detect automatically which programs are viruses and which are not by submitting them to a filter program, thus avoiding the expense and inconvenience of unwittingly and possibly harmfully altering our operating system. We now show there can be no program that does this correctly for every possible input, while guaranteed not to spread a virus itself.

We begin by fixing an operating system OS, and making a definition.

particularly appropriate to point out that one advantage of approximation via iteration is that the accuracy of the approximation is limited only by the arithmetic of the machine doing the work. This is in marked contrast to polynomial or rational approximations, for example, which must usually be completely reconstructed in order to improve accuracy.

REFERENCES

1. Kendall E. Atkinson, An Introduction to Numerical Analysis, Wiley, New York, 1978.
2. J. Hart, et al., Computer Approximations, Wiley, New York, 1968.
3. Fortran Version 5 Common Library Mathematical Routines Reference Manual, Control Data Corporation, 1979.
4. Walter Gander, On Halley's iteration method, this MONTHLY (92) 131–134.
5. George H. Brown, Jr., On Halley's variation of Newton's method, this MONTHLY (84) 726–728.
6. J. M. Borwein and P. B. Borwein, The arithmetic-geometric mean and fast computation of elementary functions, *SIAM Rev.*, 26, 351–366.
7. ______, Pi and the AGM, Wiley-Interscience, New York, 1987.

# There Are No Safe Virus Tests

WILLIAM F. DOWLING
*Department of Mathematics and Computer Science, Drexel University, Philadelphia, PA 19104*

This note gives a proof that no program can both test its input for the presence of a virus and simultaneously be guaranteed not to spread a virus itself. (You may define "virus" any way you please, as long as the definition is extensional.) This immediate corollary of Rice's Theorem [**1**] is proved by a direct diagonalization and offered as an antidote (not a vaccine) to boredom in the elementary computability course during the presentation of the halting problem.

Programs running on modern computers, unlike the executions of Turing machine programs, as usually conceived, run "in an environment." This is to say that when a program is executed, it is run as a subprogram of the logically independent program, the operating system, which is responsible for such bookkeeping chores as primary and secondary memory management, process management, recording statistics, and so on. A program that, when run, alters the code of the operating system, is called a *virus*. (When a new version of the operating system is written legitimately, it replaces, not alters, the former operating system.) This is a somewhat less restrictive definition of virus than others have proposed [**2**], in that no particular behavior is required of the modified operating system. For instance, it is frequently required that a virus have the effect of inserting its own code into other executable programs. Such restrictions are unnecessary for the result we seek.

It would be nice if we could detect automatically which programs are viruses and which are not by submitting them to a filter program, thus avoiding the expense and inconvenience of unwittingly and possibly harmfully altering our operating system. We now show there can be no program that does this correctly for every possible input, while guaranteed not to spread a virus itself.

We begin by fixing an operating system OS, and making a definition.

DEFINITION 1. *Program P spreads a virus on input x if running P under operating system OS on input x alters OS. Otherwise it is safe on input x. A program is safe if it is safe for all inputs.*

We also make the assumption that there exist viruses for OS, otherwise there would be no necessity for our test. Now for the sake of contradiction, let us assume there is some safe program *IS-SAFE* that decides the safety of running an arbitrary program $P$ on arbitrary input $x$. Thus

$$\mathit{IS\text{-}SAFE}(P, x) = \begin{cases} \text{yes} & \text{if } P \text{ is safe on input } x \\ \text{no} & \text{otherwise.} \end{cases}$$

Given such a program and our assumption that there exist viruses, it is easy to write a program $D(\,)$ of one argument that has the following behavior:

$$D(P) = \begin{cases} \text{Write "Have a nice day"} & \text{if } \mathit{IS\text{-}SAFE}(P, P) = \text{no} \\ \text{alter OS} & \text{otherwise.} \end{cases}$$

We can now show that *IS-SAFE* cannot be both safe and correct by examining the behavior of $D$ on input $D$. If $D$ is safe on input $D$ this can only be because it has not executed the otherwise clause, that is, because $\mathit{IS\text{-}SAFE}(D, D) =$ "no," thus showing that *IS-SAFE* is not correct. On the other hand, if $D$ alters OS on input $D$, there are two possibilities. On one hand, the call to $\mathit{IS\text{-}SAFE}(D, D)$ may be returning "yes" so the otherwise clause in $D$ is being executed, in which case *IS-SAFE* is not correct. On the other hand, if *IS-SAFE* returns "no" (so $D$ simply prints "Have a nice day") the assumption that $D$ is unsafe on input $D$ means that the call to *IS-SAFE* must be the culprit, that is, *IS-SAFE* is not safe. We conclude that the assumption of the existence of a safe, correct program that runs on *OS* and checks the safety of its input must be incorrect; there can be no such program *IS-SAFE*.

REFERENCES

1. H. Rogers, Theory of Recursive Functions and Effective Computability, McGraw-Hill, New York, 1967.
2. I. Witten, Computer (in)security: infiltrating open systems, *Abacus*, 4 (Summer 1987) 7–25.

## PROBLEMS AND SOLUTIONS



*For instructions about submitting **proposed** problems for publication in this department see the inside front cover. Please include solutions, relevant references, etc. Three copies are requested.*

*An asterisk (*) indicates that neither the proposer nor the editors supplied a solution.*

***Solutions** should be sent to the address given on the inside front cover. Two copies suffice.*

*A publishable solution must, above all, be correct. Given correctness, elegance and conciseness are preferred. The answer to the problem should appear right at the beginning. If your method yields a more general result, so much the better. If you discover that a Monthly problem has already been solved in the literature, you should of course tell the editors; include a copy of the solution if you can.*

*For instructions about submitting solutions of Problems, which should be mailed before January 31, 1990, see the inside front cover. Please place the solver's name and mailing address on each (doubled-spaced) sheet. Include a self-addressed card or label if an acknowledgement is desired.*

### ELEMENTARY PROBLEMS

E 3349. *Proposed by László A. Székely, University of New Mexico, Albuquerque.*

Consider the set of matchings $M$ on $2n$ labeled vertices and the set $T$ of rooted binary trees with $n+1$ labeled leaves (i.e., having $n$ unlabeled nonleaf vertices and $2n$ edges). It is well known that $|M| = (2n)!/(n!2^n)$. Prove that $T$ has the same size by giving a constructive bijection between $M$ and $T$.

E 3350. *Proposed by Paul Erdös, Hungarian Academy of Science, Budapest.*

For positive integers $n$ and $k$ with $1 \leqslant k \leqslant n$ let $A(n,k)$ denote the least common multiple of $n, n-1, \dots, n-k+1$.

(a) Prove that 14 is the largest value of $n$ for which there exists a positive integer $k \leqslant n$ such that

$$A(n,1) < A(n,2) < \cdots < A(n,k) = A(n,n),$$

i.e., 14 is the largest value of $n$ for which $A(n, j)$ is strictly increasing in $j$ until it reaches its maximum value $A(n,n)$.

(b) Let $f(n)$ be the largest value of $k$ such that

$$A(n,1) < A(n,2) < \cdots < A(n,k).$$

Prove that $f(n) < C\sqrt{n}$ for some constant $C$.

(c) Prove that $f(n)$ tends to infinity with $n$. More precisely, prove that if $n > k! + k$, then $f(n) > k$.

E 3351. *Proposed by Jet Wimp and Um Bencharit, Drexel University, Philadelphia, PA.*

Suppose $a > 0$ and let $f$ be the real-valued function on $(1, \infty)$ defined by

$$f(x) = \{1 + \operatorname{erf}(ax^{-1/2})\}^x + \{1 - \operatorname{erf}(ax^{-1/2})\}^x,$$

where

$$\operatorname{erf} u = 2\pi^{-1/2}\int_0^u e^{-t^2}\,dt.$$

Show that $f$ is monotonic on $(1, \infty)$.

E 3352. *Proposed by G. I. Senum, Brookhaven National Laboratory, Upton, Long Island, NY.*

Show that

$$\sum_{n=0}^{\infty} \frac{1}{n!(n^4 + n^2 + 1)} = \frac{e}{2}.$$

E 3354. *Proposed by V. A. Alexandrov and N. S. Dairbekov, Institute of Mathematics, Novosibirsk, USSR.*

Suppose $M:[1, \infty) \to (e, \infty)$ is a nondecreasing function such that

$$\int_1^{\infty} \frac{dt}{M(t)} = \infty. \tag{$*$}$$

(a) Prove that

$$\int_1^{\infty} \frac{dt}{t \log M(t)} = \infty.$$

(b) Show that there exist nondecreasing functions $M$ satisfying $(*)$ such that

$$\int_1^{\infty} \frac{dt}{t \log M(t) \log\log M(t)} < \infty.$$

E 3353. *Proposed by H. G. Diamond, University of Illinois at Urbana-Champaign.*

For each positive integer $n$ let $G(n)$ be the number of digits in the decimal expansion of $2^n$. Is it true that

$$\sum_{n=1}^{\infty} \frac{G(n)}{2^n} = \frac{1169}{1023}?$$

(Cf. Problem 6609 in the previous issue.)

## SOLUTIONS OF ELEMENTARY PROBLEMS

### Uniqueness of a Boundary Value Problem

E 3218 [1987, 549]. *Proposed by W. O. Egerland, University of Baltimore, Baltimore, MD.*

Suppose $a$ and $T$ are given positive real numbers, and suppose $n$ is a given positive integer. Assume that the differential equation $x^{(2n)} = (-1)^n ax$ with bound-

*Editorial comment.* The simplest examples of such polynomials (for which the zeros have no common factor) are as follows:

$$x^3 - 147x + 286 = (x-2)(x-11)(x+13)$$
$$x^3 - 507x + 506 = (x-1)(x-22)(x+23)$$
$$x^3 - 1083x + 10582 = (x-11)(x-26)(x+37)$$
$$x^3 - 2883x + 34282 = (x-13)(x-46)(x+59)$$
$$x^3 - 4107x + 89206 = (x-26)(x-47)(x+73)$$
$$x^3 - 5547x + 111386 = (x-22)(x-61)(x+83)$$
$$x^3 - 7203x + 98098 = (x-14)(x-77)(x+91)$$
$$x^3 - 7203x + 153502 = (x-23)(x-71)(x+94).$$

Several solvers reduced the problem to a consideration of the diophantine equation $u^2 + 3v^2 = k$. In fact, without loss of generality we may assume that $k$ is odd (since, if $k$ is even, (1) shows that $a, b, c$ are all even). With $k$ odd exactly one of $a, b, c$ is even. Without loss of generality we may suppose that $b$ is even, say $b = 2u$. Then (2) gives $(a+u)^2 + 3u^2 = 3k^2$, so that $a + u$ is divisible by 3, say $a + u = 3v$. Then $u^2 + 3v^2 = k^2$ and we may determine $a, b$ from $u, v$ by the equalities $a = 3v - u$, $b = 2u$.

Related results dealing with the characterization of polynomials having integral roots and extrema are:

Tom Bruggeman and Tom Gush, Nice cubic polynomials for curve sketching, *Math. Mag.*, 53 (1980) 233–234.
Frank Schmidt, Problem 87-9, *Math. Intelligencer* 9, no. 3 (1987) 40.
Johann Walter, Über ganze rationale Funktionen dritten Grades mit ganzzahligen Koeffizienten, bei denen Nullstellen and Extrema zugleich ganzzahlig sind, *Praxis der Mathematik*, 29 (1987) 489–492.

Also solved by J. C. Binz (Switzerland), K. Brown, J. Delany, F. Dodd, W. Janous (Austria), L. E. Mattics, J. B. Muskat and W. Saffern (Israel), A. Pedersen (Denmark), S. Philipp, V. Schindler (East Germany), J. H. Steelman, M. Vowe (Switzerland), J. Walter (West Germany), and the proposer. Several incorrect or incomplete solutions were received.

### Relatively Sparse Subsequences

E 3268 [1988, 457]. *Proposed by Dorothy Maharam and A. H. Stone, University of Rochester, Rochester, NY.*

Does there exist a sequence $\xi_1, \xi_2, \ldots$ of positive real numbers such that whenever $0 < \alpha < \beta$, we have

$$\lim_{k\to\infty} \frac{\#\{n : 1 \leqslant n \leqslant k,\ \xi_n > \beta\}}{\#\{n : 1 \leqslant n \leqslant k,\ \xi_n > \alpha\}} = 0?$$

*Solution by Nicholas Passell, University of Wisconsin, Eau Claire, WI.* Yes. Let $\{r_i : 1 \leqslant i < \infty\}$ be any enumeration of the positive rational numbers and let $S = \{s_{m,j} : 1 \leqslant j \leqslant m,\ 1 \leqslant m < \infty\}$ be the lower triangular array whose $m$th row consists of the set $\{r_i : 1 \leqslant i \leqslant m\}$ rearranged in increasing order. We construct the

sequence $\xi_1, \xi_2, \ldots$ by starting with $s_{1,1}$, then taking $s_{2,1}$ twice and $s_{2,2}$ once, and proceeding through each subsequent row of $S$ in turn from left to right, taking $s_{m,j}$ exactly $m^{m-j}$ times. We claim that the sequence so constructed has the property stated in the problem.

By the $m$th block of the sequence $\{\xi_n\}$ we mean the $m^{m-1} + m^{m-2} + \cdots + m + 1 = (m^m - 1)/(m-1)$ terms so obtained from the $m$th row of $S$. Of the $m$ elements $s_{m,j}$ $(1 \leqslant j \leqslant m)$ of this row suppose $a_m$ are in $(0, \alpha]$, $b_m$ are in $(\alpha, \beta]$, and $c_m$ are in $(\beta, \infty)$. By construction $a_{m+1} \geqslant a_m$, $b_{m+1} \geqslant b_m$, and $c_{m+1} \geqslant c_m$ and, since $\mathbb{Q}$ is dense in $\mathbb{R}$, each of $a_m, b_m, c_m$ tends to infinity with $m$.

We consider what happens to the ratio in the statement of the problem as we go through the $m$th block of $\{\xi_n\}$, assuming that $m$ is large enough so that $a_m, b_m, c_m$ are all positive. When $\xi_n = s_{m,j}$ with $j \leqslant a_m$, the count in both numerator and denominator is unchanged when we go from $n-1$ to $n$. When $\xi_n = s_{m,j}$ with $a_m < j \leqslant a_m + b_m$, the denominator increases by one and the numerator is unchanged when we go from $n-1$ to $n$. Finally when $\xi_n = s_{m,j}$ with $a_m + b_m < j$, both numerator and denominator increase by one when we go from $n-1$ to $n$. Thus the largest ratio occurs at the beginning or the end of the block.

Let $N_m$ denote the number of $\xi$'s in the $m$th block which are greater than $\beta$ and let $D_m$ denote the number of $\xi$'s in the $m$th block greater than $\alpha$. The argument of the preceding paragraph shows that it suffices to prove that $(N_1 + \cdots + N_m)/(D_1 + \cdots + D_m)$ approaches zero. But this follows from the limiting assertions $D_m \to \infty$ and $N_m/D_m \to 0$, which in turn follow from the explicit formulas

$$N_m = (m^{c_m} - 1)/(m-1), \qquad D_m = (m^{b_m + c_m} - 1)/(m-1)$$

and the resulting inequality

$$N_m/D_m \leqslant m^{-b_m}.$$

Also solved by K. Austin (England), S. F. Barger, R. B. Eggleton (Australia), T.-Y. Mao (China), K. A. Ross, K. Schilling, J. H. Steelman, P. Tracy, D. M. Wells, J. B. Wilker (Canada), Univ. of South Alabama Problem Group, and the proposers.

### Permutations with Distinct Displacements

E 3269 [1988, 554]. *Proposed by M. J. Pelling, University College, London, England.*

For what positive integers $n$ does there exist a permutation $(x_1, x_2, \ldots, x_n)$ of $(1, 2, \ldots, n)$ such that the differences $|x_k - k|$, $1 \leqslant k \leqslant n$, are all distinct?

*Solution by John Henry Steelman, Indiana University of Pennsylvania, Indiana, PA.* Such a permutation exists if and only if $n \equiv 0$ or $1 \pmod 4$.

To see that this condition is necessary, let $d_k = x_k - k$. Note that $0 \leqslant |d_k| \leqslant n-1$. If these differences are all distinct, then

$$\sum_{k=1}^{n} |d_k| = \sum_{j=0}^{n-1} j = \frac{n(n-1)}{2}.$$

But $\Sigma |d_k| \equiv \Sigma d_k = 0 \pmod 2$. Thus $n(n-1)/2$ must be even and hence $n \equiv 0$ or $1 \pmod 4$.

Conversely, if $n \equiv 0$ or 1 (mod 4), one can construct a permutation with the desired property by letting

$$x_k = \begin{cases} n+1-k & \text{if } \quad 1 \leqslant k \leqslant \lfloor n/4 \rfloor \\ n-k & \text{if } \quad \lfloor n/4 \rfloor < k < \lceil n/2 \rceil \\ 1 & \text{if } \quad k = \lceil n/2 \rceil \\ n+1-k & \text{if } \quad \lceil n/2 \rceil < k < \lceil 3n/4 \rceil \\ k & \text{if } \quad k = \lceil 3n/4 \rceil \\ n+2-k & \text{if } \quad \lceil 3n/4 \rceil < k \leqslant n \end{cases}$$

Here $\lfloor x \rfloor$ denotes the greatest integer less than or equal to $x$ and $\lceil x \rceil$ denotes the least integer greater than or equal to $x$.

Also solved by the proposer and 22 other readers. One partial solution was received.

#### Rationality of a Trigonometric Function

E 3270 [1988, 554]. *Proposed by Berndt Lindström, University of Stockholm, Sweden.*

Determine those positive rational numbers $m$ such that

$$\frac{1}{\pi}\arctan\sqrt{m}$$

is also rational.

*Solution I by Diane Johnson and Roy Dowling, University of Manitoba, Winnipeg, Manitoba.* The numbers are $1/3$, 1, and 3. Suppose $k = (1/\pi)\arctan\sqrt{m}$ where $k$ is rational and $m$ is positive and rational. Then

$$m = \tan^2 k\pi = \frac{1 - \cos 2k\pi}{1 + \cos 2k\pi}.$$

Thus $\cos 2k\pi = (1-m)/(1+m)$ is rational. However, it is known that the only rational values of $\cos 2r\pi$ with $r$ rational are 0, $\pm\frac{1}{2}$, and $\pm 1$. (See I. Niven, *Irrational Numbers*, Carus Monograph No. 11, John Wiley, 1956, page 41, or J. M. H. Olmsted, *Amer. Math. Monthly*, 52 (1945) 336–337.) Since $(1-m)/(1+m)$ cannot equal $-1$ and $m$ must be positive, it follows that the possible values of $m$ are $1/3$, 1, and 3.

*Solution II by Duane Broline, Eastern Illinois University, Charleston, IL.* If $\sqrt{m} = \tan r\pi$, where $r$ and $m$ are rational, then

$$2\cos 2r\pi = 2\left(\frac{2}{\sec^2 r\pi} - 1\right) = \frac{4}{m+1} - 2.$$

Since $2\cos 2r\pi = e^{2r\pi i} + e^{-2r\pi i}$ is both an algebraic integer and a rational number, it must be an integer at most 2 in absolute value. Thus $4/(m+1)$ must be an integer between 0 and 4 inclusive. The values 0 and 4 are impossible, while 1, 2, and 3 yield $m = 3$, 1, and $1/3$, respectively.

*Editorial comment.* David W. Koster noted that this problem appears as problem 197.5 of part VIII of *Problems and Theorems in Analysis II*, by G. Pólya and G. Szegö, Springer-Verlag, 1972, 1978. (It does not appear in the original German edition.)

Also solved by M. Divovich, J. Ferrer, S. M. Gagola, Jr., D. Hildebrandt, M. Khan (student), K.-W. Lau (Hong Kong), O. P. Lossers (The Netherlands), D. E. Manes, A. Nijenhuis, J. Rebholz (student), U. of South Alabama Problem Group, and the proposer.

### A Sum Involving the Zeta Function

E 3271 [1988, 554]. *Proposed by Walter Noll and Joseph C. Keane, Carnegie Mellon University, Pittsburgh, PA.*

For each $r \in \mathbb{Q}$ (the set of all rational numbers), denote the denominator of the representation of $r$ as a reduced fraction by $\delta(r)$. For each $\alpha \in \mathbb{R}$ (the set of all real numbers) consider

$$\sigma(\alpha) := \sum_{r \in S} (\delta(r))^{-\alpha},$$

where

$$S := \{r \in \mathbb{Q} : 0 < r \leqslant 1\}.$$

Show that $\sigma(\alpha)$ is finite if and only if $\alpha > 2$, and evaluate $\sigma(\alpha)$ when $\alpha > 2$.

*Solution by Stephen M. Gagola, Jr., Kent State University, Kent, OH.* Since there are exactly $\phi(n)$ reduced fractions in $S$ having denominator $n$, we have $\sigma(\alpha) = \Sigma_{n \geqslant 1}\phi(n)n^{-\alpha}$. Because $\phi(n)n^{-\alpha} = (\phi(n)/n)n^{-(\alpha-1)} \leqslant n^{-(\alpha-1)}$, the series converges for $\alpha > 2$. If $\alpha \leqslant 2$ and $p$ is a prime, then the $p$th term of this series satisfies $(p-1)p^{-\alpha} = (1 - 1/p)p^{1-\alpha} \geqslant p^{-1}/2$. Since the sum of $1/p$ over all primes $p$ diverges, the series for $\sigma(\alpha)$ diverges if $\alpha \leqslant 2$.

Fix $\alpha > 2$ and multiply the series for $\sigma(\alpha)$ by $\zeta(\alpha) = \Sigma_{n \geqslant 1} n^{-\alpha}$ to obtain $\Sigma_{n \geqslant 1}(\Sigma_{d|n}\phi(d))n^{-\alpha} = \Sigma_{n \geqslant 1} n \cdot n^{-\alpha} = \zeta(\alpha - 1)$. Hence $\sigma(\alpha) = \zeta(\alpha-1)/\zeta(\alpha)$.

*Editorial comment.* Readers provided several relevant references where the problem or solution more or less explicitly appears: G. H. Hardy and E. M. Wright, *The Theory of Numbers*, Theorem 288; M. Abramovitz and I. A. Stegun, *Handbook of Mathematical Functions*, p. 826; J. B. Conway, *Functions of One Complex Variable*, p. 190, problem 5.

Also solved by M. Berg & D. Tyler, D. Callan, J. Dalbec, J. Deutsch, D. Doster, N. Franceschine, J. Guillera (Spain), W. H. Gustafson, M. V. Hegde, W. Janous (Austria), H.-S. Ki (student, South Korea), I. E. Leonard (Canada), O. P. Lossers (The Netherlands), T. McDonald, A. Shimron (Israel), C. Schoen, K. Zacharias (West Germany), Brown Univ. Fly-Fishing Club, SUNY $\Pi_1$, Univ. of S. Alabama Problem Group, and the proposers.

### Operation Count for a Determinant Computation

E 3272 [1988, 555]. *Proposed by K. I. Appel and C. G. Jockusch, Jr., University of Illinois at Urbana-Champaign.*

It is well known that $(1/3)(n^3 + 2n - 3)$ multiplications/divisions are required to find the determinant of an $n \times n$ matrix by Gaussian elimination, while obvi-

ously the permutation expansion requires $(n-1)n!$ multiplications. Obtain a formula in closed form for the number $a_n$ of multiplications needed to evaluate an $n \times n$ determinant by the cofactor method. The formula should not involve summations but may use the greatest integer function.

*Solution by G. W. Valk, Tucker, GA.* The answer is $a_n = \lfloor n!(e-1) - 1 \rfloor$. The recurrence relation for $a_n$ is $a_n = n(a_{n-1} + 1)$. It is easy to verify by substitution that $a_n = n!\Sigma_{i=1}^{n-1} 1/i!$ satisfies this and the initial condition $a_2 = 2$. Hence $a_n = n!(e-1) - 1 - R_{n+1}$, where $R_{n+1} = n!\Sigma_{i=n+1}^{\infty} 1/i!$. Since $0 < R_{n+1} < \Sigma_{i=1}^{\infty} 1/n^i$ and $\Sigma_{i=1}^{\infty} 1/n^i = 1/(n-1)$ if $1/n < 1$, we have $0 < R_{n+1} < 1/(n-1) \leqslant 1$. Since $a_n$ is an integer, we have $a_n = \lfloor n!(e-1) - 1 \rfloor$.

*Editorial comment.* This problem is also solved in the paper, "A difference equation and operation counts in computation of determinants," *Mathematics Magazine*, 61 (1988) 295–297, by S. H. Friedberg.

Also solved by D. Callan, P.-C. Chuang, J. Duemmel, J. Foster, T. McDonald, T. S. Norfolk, A. Pedersen (Denmark), L. R. Purser, J. H. Steelman, The Brown University Fly-Fishing Club, and the proposers.

## ADVANCED PROBLEMS

6612. *Proposed by Ebrahim Salehi, University of Nevada, Las Vegas.*

Suppose $X$ is a compact metric space with metric $d$ and suppose $T: X \to X$ is continuous. If

$$\inf_{n \in \mathbb{N}} d(T^n x, T^n y) > 0$$

for each pair $x$, $y$ of distinct elements of $X$, prove that $T$ is onto.

6613. *Proposed by Jeff Vaaler, University of Texas, Austin, TX.*

Suppose $P$ is a polynomial of degree $n$ with complex coefficients, leading coefficient $a$, and zeros $\alpha_1, \dots, \alpha_n$. Define the Mahler measure $M(P)$ of $P$ by

$$M(P) = \exp \int_0^1 \log|P(e^{2\pi i t})|\, dt = a \prod_{j=1}^{n} \max(1, |\alpha_j|),$$

where the latter equality follows by Jensen's formula. If $P'$ is the derivative of $P$, prove that

$$M(P') \leqslant nM(P).$$

6614. *Proposed by Justin G. MacCarthy, Deming, NM.*

A nonempty proper subset $A$ of an infinite abelian group $G$ is called a "wedge" in $G$ if the family of all translates $A + x$, $x \in G$, is closed under finite intersection. Prove that in $Q^+$, the additive group of the rationals, the only wedges are half-lines (open or closed).

6615. *Proposed by Kee-Wai Lau, Hong Kong.*

Denote by $(i, j)$ and $[i, j]$ respectively the greatest common divisor and the least common multiple of the positive integers $i$ and $j$. Prove that as $n$ tends to infinity

$$\text{(i)} \qquad \sum_{i=1}^{n} \sum_{j=1}^{n} (i, j) \sim \frac{n^2 \ln n}{\zeta(2)},$$

$$\text{(ii)} \qquad \sum_{i=1}^{n} \sum_{j=1}^{n} [i, j] \sim \frac{\zeta(3) n^4}{4\zeta(2)},$$

where $\zeta$ denotes the Riemann zeta function.

## SOLUTIONS OF ADVANCED PROBLEMS

### Cubical Walks

6556 [1987, 800]. *Proposed by N. J. Fine, Deerfield Beach, Florida.*

(a) Consider a random walk around the edges of a square, where the probability of moving from a given vertex to either of the two adjacent vertices is $1/2$. Suppose the walk stops as soon as all edges have been traversed. Find the expected path-length.

*(b) Consider a random walk around the edges of a cube, where the probability of moving from a given vertex to any one of the three adjacent vertices is $1/3$. Find the expected path-length needed to traverse all edges.

*(c) Similarly with the frame of the $n$-dimensional cube, where each probability is $1/n$.

*Solution to parts (a) and (b) by Peter A. Griffin, California State University, Sacramento.* (a) The expected path length (number of steps) to traverse all the edges of a square is $X = 10$. It is found by solving the system (suggested by the accompanying diagrams):

$$A = 1 + B/2$$

$$B = 1 + A/2 + B/2$$

$$C = 1 + D/2 + A/2$$

$$D = 1 + C$$

$$E = 1 + E/2 + C/2$$

$$X = 1 + E$$

*A* *B* *C* *D* *E*

6615. *Proposed by Kee-Wai Lau, Hong Kong.*

Denote by $(i, j)$ and $[i, j]$ respectively the greatest common divisor and the least common multiple of the positive integers $i$ and $j$. Prove that as $n$ tends to infinity

$$\text{(i)} \qquad \sum_{i=1}^{n} \sum_{j=1}^{n} (i, j) \sim \frac{n^2 \ln n}{\zeta(2)},$$

$$\text{(ii)} \qquad \sum_{i=1}^{n} \sum_{j=1}^{n} [i, j] \sim \frac{\zeta(3) n^4}{4\zeta(2)},$$

where $\zeta$ denotes the Riemann zeta function.

## SOLUTIONS OF ADVANCED PROBLEMS

### Cubical Walks

6556 [1987, 800]. *Proposed by N. J. Fine, Deerfield Beach, Florida.*

(a) Consider a random walk around the edges of a square, where the probability of moving from a given vertex to either of the two adjacent vertices is $1/2$. Suppose the walk stops as soon as all edges have been traversed. Find the expected path-length.

*(b) Consider a random walk around the edges of a cube, where the probability of moving from a given vertex to any one of the three adjacent vertices is $1/3$. Find the expected path-length needed to traverse all edges.

*(c) Similarly with the frame of the $n$-dimensional cube, where each probability is $1/n$.

*Solution to parts (a) and (b) by Peter A. Griffin, California State University, Sacramento.* (a) The expected path length (number of steps) to traverse all the edges of a square is $X = 10$. It is found by solving the system (suggested by the accompanying diagrams):

$$\begin{aligned} A &= 1 + B/2 \\ B &= 1 + A/2 + B/2 \\ C &= 1 + D/2 + A/2 \\ D &= 1 + C \\ E &= 1 + E/2 + C/2 \\ X &= 1 + E \end{aligned}$$

$A$ $\qquad$ $B$ $\qquad$ $C$ $\qquad$ $D$ $\qquad$ $E$

walk induces a Markov chain on this state space. If pairs of directed edges containing the same vertices are considered as sets, the original problem becomes: how long does it take this Markov chain to visit all of these sets? Aldous [On the time taken by random walks on finite groups to visit every state, *Z. Wahrsch. Verw. Gebiete* 62 (1983) 361–374] gives heuristics applicable to this problem." These heuristics suggest that, if $T$ is the time required to visit all pairs, then

$$E(T) = n2^{n-1} \cdot n \log 2 \cdot (1 + o(1)).$$

Mathews remarks that this can be made rigorous using bounds he has given [Covering problems for Brownian motion on spheres, *Ann. Prob.* 16 (1988) 189–199].

Thus, asymptotically, the result is the same as for the coupon problem, namely, (number of edges) · log(number of edges). Mathews gives an argument to suggest the asymptotically equivalent but more accurate approximation

$$n2^{n-1}\left(\frac{n}{n-1}\right)\{\log(n2^{n-1}) + \gamma\}.$$

This gives 55.12 for $n = 3$. Here $\gamma$ is Euler's constant.

A proof that the expected number of times for every *vertex* to be visited is asymptotically number of vertices · log(number of vertices) = $2^n \log 2^n$ follows from the results of Aldous cited above or Mathews, P.C., Covering problems for Markov chains, *Ann. Prob.* 16 (1988) 1215–1228. Thus, again the asymptotic result is the same as for the independent model. For $n = 2$ the expected number is 6, while for $n = 3$ it is 1996/95.

The problem of obtaining upper and lower bounds for the expected number of steps for a random walk on a graph to reach every state is important in computer science applications and is the subject of a great deal of current research. In work under preparation Chandra, Raghavan, Ruzzo, Smolensky, and Tiwari show that bounds can be expressed in terms of the electrical resistance of the graph as an electric network with unit resistors as edges.

An extensive bibliography of the work on covering theorems and random walks can be obtained from David J. Aldous, Department of Statistics, University of California, Berkeley.

The editors wish to thank Professor J. L. Snell of Dartmouth College for his assistance in preparing these comments. Professor Snell is the coauthor of *Random Walks and Electric Networks*, Carus Mathematical Monographs, No. 22.

### Beta Function Determinants

6561 [1987, 1011]. *Proposed by Heinz-Jürgen Seiffert, Berlin.*

For complex numbers $z$ and $w$ with $\mathrm{Re}(z) > 0$ and $\mathrm{Re}(w) > 0$ consider the matrices

$$C_n(z, w) := \left(B(z + j + k, w)\right)_{j,k=0,\dots,n},$$

$$D_n(z, w) := \left(\frac{1}{B(z + j + k, w)}\right)_{j,k=0,\dots,n},$$

walk induces a Markov chain on this state space. If pairs of directed edges containing the same vertices are considered as sets, the original problem becomes: how long does it take this Markov chain to visit all of these sets? Aldous [On the time taken by random walks on finite groups to visit every state, *Z. Wahrsch. Verw. Gebiete* 62 (1983) 361–374] gives heuristics applicable to this problem." These heuristics suggest that, if $T$ is the time required to visit all pairs, then

$$E(T) = n2^{n-1} \cdot n \log 2 \cdot (1 + o(1)).$$

Mathews remarks that this can be made rigorous using bounds he has given [Covering problems for Brownian motion on spheres, *Ann. Prob.* 16 (1988) 189–199].

Thus, asymptotically, the result is the same as for the coupon problem, namely, (number of edges) $\cdot$ log(number of edges). Mathews gives an argument to suggest the asymptotically equivalent but more accurate approximation

$$n2^{n-1}\left(\frac{n}{n-1}\right)\{\log(n2^{n-1}) + \gamma\}.$$

This gives 55.12 for $n = 3$. Here $\gamma$ is Euler's constant.

A proof that the expected number of times for every *vertex* to be visited is asymptotically number of vertices $\cdot$ log(number of vertices) $= 2^n \log 2^n$ follows from the results of Aldous cited above or Mathews, P.C., Covering problems for Markov chains, *Ann. Prob.* 16 (1988) 1215–1228. Thus, again the asymptotic result is the same as for the independent model. For $n = 2$ the expected number is 6, while for $n = 3$ it is 1996/95.

The problem of obtaining upper and lower bounds for the expected number of steps for a random walk on a graph to reach every state is important in computer science applications and is the subject of a great deal of current research. In work under preparation Chandra, Raghavan, Ruzzo, Smolensky, and Tiwari show that bounds can be expressed in terms of the electrical resistance of the graph as an electric network with unit resistors as edges.

An extensive bibliography of the work on covering theorems and random walks can be obtained from David J. Aldous, Department of Statistics, University of California, Berkeley.

The editors wish to thank Professor J. L. Snell of Dartmouth College for his assistance in preparing these comments. Professor Snell is the coauthor of *Random Walks and Electric Networks*, Carus Mathematical Monographs, No. 22.

### Beta Function Determinants

6561 [1987, 1011]. *Proposed by Heinz-Jürgen Seiffert, Berlin.*

For complex numbers $z$ and $w$ with $\mathrm{Re}(z) > 0$ and $\mathrm{Re}(w) > 0$ consider the matrices

$$C_n(z, w) := \left(B(z + j + k, w)\right)_{j,k=0,\dots,n},$$

$$D_n(z, w) := \left(\frac{1}{B(z + j + k, w)}\right)_{j,k=0,\dots,n},$$

where $B$ denotes the beta function

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} = \int_0^1 t^{x-1}(1-t)^{y-1}\,dt.$$

Show that

(a) $\det C_n(z, w) = nB(z+w+n-1, n)B(z+n, w+n)\det C_{n-1}(z, w)$,

(b) $\det D_n(z, w) = q_n(w)\dfrac{B(z+n-1, n)}{B(z+n, w)}\dfrac{\Gamma(z+n)}{\Gamma(z+2n)}\det D_{n-1}(z, w)$,

where

$$q_n(w) = n\prod_{k=0}^{n-1}(k-w).$$

*Solution by the proposer.* Both sides of (a) are analytic in $z$ and $w$. Hence it suffices to show equality when $z = a$ and $w = b$, where $a$ and $b$ are real and positive. Define an inner product by

$$(f, g) := \int_0^1 f(t)g(t)t^{a-1}(1-t)^{b-1}\,dt,$$

and let $J_0, J_1, J_2, \ldots$ be the sequence of orthonormal polynomials derived from the powers $1, t, t^2, \ldots,$ by the Gram-Schmidt process. These are special Jacobi polynomials, and $J_n$ has leading coefficient

$$c_n = \left[nB(a+b+n-1, n)B(a+n, b+n)\right]^{-1/2}$$

(see F. G. Tricomi, *Vorlesungen über Orthogonalreihen,* Springer Verlag, Berlin, 1970, pp. 160–163). Define $f_j(t) = t^j$ for $j = 0, \ldots, n$, and $f_{n+1}(t) = J_n(t)$. Then

$$(f_j, f_{n+1}) = 0, \qquad j = 0, \ldots, n-1,$$

$$(f_n, f_{n+1}) = \frac{1}{c_n},$$

and

$$(f_{n+1}, f_{n+1}) = 1.$$

Now $f_0, \ldots, f_{n+1}$ are linearly dependent, so their Gram-determinant $G(f_0, \ldots, f_{n+1})$ vanishes. This means that

$$0 = G(f_0, \ldots, f_{n+1}) = \left|\begin{array}{cc|c} \multicolumn{2}{c|}{C_n(a, b)} & \begin{array}{c} O_n \\ \hline \dfrac{1}{c_n} \end{array} \\ \hline O_n' & \dfrac{1}{c_n} & 1 \end{array}\right|$$

$$= \det C_n(a, b) - \frac{1}{c_n^2}\det C_{n-1}(a, b),$$

where $O_n$ is a column of $n$ zeros and $O_n'$ is a row of $n$ zeros. This proves (a).

For

$$\tilde{\Gamma}_n(z) := \left(\frac{1}{\Gamma(z+j+k)}\right)_{j,k=0,1,\ldots,n}$$

we have

$$\det \tilde{\Gamma}_n(z) = (-1)^n \frac{(n-1)!\Gamma(z+n-1)}{\Gamma(z+2n-1)\Gamma(z+2n)} \det \tilde{\Gamma}_{n-1}(z).$$

*Editorial comment.* The results mentioned in the last paragraph can be obtained from the assertions of the problem by noting that $B(x,\lambda) \sim \lambda^{-x}\Gamma(x)$ as $\lambda \to \infty$, so that as $\lambda \to \infty$ we have

$$\det C_n(z,\lambda) \sim \lambda^{-(n+1)z-n^2-n} \det \Gamma_n(z)$$

and

$$\det D_n(z,\lambda) \sim \lambda^{(n+1)z+n^2+n} \det \tilde{\Gamma}_n(z).$$

No other solutions were received.

**Between Euler and Jacobi**

6562 [1987, 1011]. *Proposed by George E. Andrews, Pennsylvania State University, University Park.*

Let $Q(q) = \prod_{n=1}^{\infty}(1-q^n)$ for $|q|<1$. Euler's Pentagonal Number Theorem asserts that

$$Q(q) = \sum_{n=0}^{\infty} (-1)^n q^{n(3n+1)/2}(1-q^{2n+1}).$$

Jacobi showed that

$$Q(q)^3 = \sum_{n=0}^{\infty} (-1)^n (2n+1) q^{n(n+1)/2}.$$

Prove that

$$Q(q)^2 = \sum_{n=0}^{\infty} (-1)^n q^{n(n+1)/2}(1-q^{2n+2}) p_n(q),$$

where $p_n(q) = \sum_{r=0}^{n} q^{r(n-r)}$.

*Solution by the Proposer.* For the solution we require two fairly elementary facts: The $q$-binomial theorem [**1**; p. 36, eq. (3.3.6)]

$$\sum_{j=0}^{N} \frac{(q)_N}{(q)_j (q)_{N-j}} (-1)^j z^j q^{j(j-1)/2} = (z)_N, \tag{1}$$

and a special case of Heine's transformation [**1**; p. 19, Cor. 2.3, $a=b=0$, $t=q$,

For

$$\tilde{\Gamma}_n(z) := \left(\frac{1}{\Gamma(z+j+k)}\right)_{j,k=0,1,\ldots,n}$$

we have

$$\det \tilde{\Gamma}_n(z) = (-1)^n \frac{(n-1)!\Gamma(z+n-1)}{\Gamma(z+2n-1)\Gamma(z+2n)} \det \tilde{\Gamma}_{n-1}(z).$$

*Editorial comment.* The results mentioned in the last paragraph can be obtained from the assertions of the problem by noting that $B(x, \lambda) \sim \lambda^{-x}\Gamma(x)$ as $\lambda \to \infty$, so that as $\lambda \to \infty$ we have

$$\det C_n(z, \lambda) \sim \lambda^{-(n+1)z-n^2-n} \det \Gamma_n(z)$$

and

$$\det D_n(z, \lambda) \sim \lambda^{(n+1)z+n^2+n} \det \tilde{\Gamma}_n(z).$$

No other solutions were received.

**Between Euler and Jacobi**

6562 [1987, 1011]. *Proposed by George E. Andrews, Pennsylvania State University, University Park.*

Let $Q(q) = \prod_{n=1}^{\infty}(1-q^n)$ for $|q| < 1$. Euler's Pentagonal Number Theorem asserts that

$$Q(q) = \sum_{n=0}^{\infty} (-1)^n q^{n(3n+1)/2}(1-q^{2n+1}).$$

Jacobi showed that

$$Q(q)^3 = \sum_{n=0}^{\infty} (-1)^n (2n+1) q^{n(n+1)/2}.$$

Prove that

$$Q(q)^2 = \sum_{n=0}^{\infty} (-1)^n q^{n(n+1)/2}(1-q^{2n+2}) p_n(q),$$

where $p_n(q) = \sum_{r=0}^{n} q^{r(n-r)}$.

*Solution by the Proposer.* For the solution we require two fairly elementary facts: The $q$-binomial theorem [**1**; p. 36, eq. (3.3.6)]

$$\sum_{j=0}^{N} \frac{(q)_N}{(q)_j(q)_{N-j}} (-1)^j z^j q^{j(j-1)/2} = (z)_N, \tag{1}$$

and a special case of Heine's transformation [**1**; p. 19, Cor. 2.3, $a = b = 0$, $t = q$,

$c = q^{2r+1}$]

$$\sum_{n=0}^{\infty} \frac{q^n}{(q)_n (q)_{n+2r}} = \frac{1}{(q)_\infty^2} \sum_{m=0}^{\infty} (-1)^m q^{m(m+1)/2+2rm}, \qquad |q| < 1, \tag{2}$$

where

$$(A)_n = (1-A)(1-Aq)\cdots(1-Aq^{n-1}), \qquad (A)_\infty = \lim_{n\to\infty} (A)_n. \tag{3}$$

Noting that $(z)_0 \equiv 1$ and $(q^{-N})_{2N} = 0$ for positive integral $N$, we see that

$$\begin{aligned}
1 &= \sum_{N=0}^{\infty} \frac{(-1)^N (q^{-N})_{2N}}{(q)_{2N}} q^{N(N+3)/2} \\
&= \sum_{N=0}^{\infty} (-1)^N \sum_{n=0}^{2N} \frac{(-1)^n q^{n(n-1)/2 - Nn + N(N+1)/2 + N}}{(q)_n (q)_{2N-n}} \qquad \text{(by (1))} \\
&= \sum_{N=0}^{\infty} q^N \sum_{n=-N}^{N} \frac{(-1)^n q^{n(n-1)/2}}{(q)_{n+N} (q)_{N-n}} \\
&= \sum_{N=0}^{\infty} q^N \left( \frac{1}{(q)_N^2} + \sum_{n=1}^{N} \frac{(-1)^n q^{n(n-1)/2}(1+q^n)}{(q)_{N+n} (q)_{N-n}} \right) \\
&= \sum_{N=0}^{\infty} q^N \sum_{n=0}^{N} \frac{\alpha_n}{(q)_{N+n} (q)_{N-n}},
\end{aligned} \tag{4}$$

where $\alpha_0 = 1$, $\alpha_n = (-1)^n q^{n(n-1)/2}(1+q^n)$ for $n > 0$.

Therefore, multiplying (4) by $(q)_\infty^2$, we obtain

$$\begin{aligned}
(q)_\infty^2 &= (q)_\infty^2 \sum_{n=0}^{\infty} \alpha_n \sum_{N=n}^{\infty} \frac{q^N}{(q)_{N+n} (q)_{N-n}} \\
&= (q)_\infty^2 \sum_{n=0}^{\infty} \alpha_n q^n \sum_{N=0}^{\infty} \frac{q^N}{(q)_N (q)_{N+2n}} \\
&= (q)_\infty^2 \sum_{n=0}^{\infty} \alpha_n q^n \frac{1}{(q)_\infty^2} \sum_{m=0}^{\infty} (-1)^m q^{m(m+1)/2+2nm} \qquad \text{(by (2))} \\
&= \sum_{m=0}^{\infty} (-1)^m q^{m(m+1)/2} \left( 1 + \sum_{n=1}^{\infty} (-1)^n q^{n(n+1)/2} (1+q^n) q^{2nm} \right) \\
&= \sum_{m=0}^{\infty} (-1)^m q^{m(m+1)/2} \sum_{n=0}^{\infty} (-1)^n q^{n(n+1)/2+2nm} \\
&\quad + \sum_{m=0}^{\infty} (-1)^m q^{m(m+1)/2} \sum_{n=1}^{\infty} (-1)^n q^{n(n+1)/2+n+2nm}
\end{aligned}$$

$$= \sum_{N=0}^{\infty} (-1)^N q^{N(N+1)/2} \sum_{\substack{m \geqq 0,\, n \geqq 0 \\ m+n=N}} q^{mn}$$

$$- \sum_{m=0}^{\infty} (-1)^m q^{m(m+1)/2} \sum_{n=0}^{\infty} (-1)^n q^{n(n+1)/2+2n+2nm+2m+2}$$

$$= \sum_{N=0}^{\infty} (-1)^N q^{N(N+1)/2} p_N(q)$$

$$- \sum_{N=0}^{\infty} (-1)^N q^{N(N+1)/2+2N+2} \sum_{\substack{m \geqq 0,\, n \geqq 0 \\ m+n=N}} q^{mn}$$

$$= \sum_{N=0}^{\infty} (-1)^N q^{N(N+1)/2} (1 - q^{2N+2}) p_N(q), \tag{5}$$

where $p_N(q) = \sum_{m=0}^{N} q^{m(N-m)}$. □

A combinatorial proof of this result has been found by J. T. Joichi [2].

No other solutions were received.

REFERENCES

1. G. E. Andrews, The Theory of Partitions, The Encyclopedia of Mathematics, Vol. 2, G.-C. Rota, Ed., Addison-Wesley, Reading, 1976 (Reprinted: Cambridge University Press, London and New York, 1985).
2. J. T. Joichi, Hecke-Rogers-Andrews identities: combinatorial proofs, *Discr. Math.* (to appear).

# REVIEWS

EDITED BY JOSEPH KONHAUSER
*Department of Mathematics, Macalester College, St. Paul, MN 55105*

*Mathematica—A System for Doing Mathematics by Computer.* Wolfram Research, Inc., Champaign, IL.

LAWRENCE S. KROLL
*Department of Computer Science, San Francisco State University, San Francisco, CA 94132*

When I was going to school in Berkeley I kept a giant 10-pound book called *The Handbook of Chemistry and Physics* at my desk. Every time I came to an unknown integral or wanted a formula I would look for it in this tome. It seemed to hold everything I needed to know about my courses. Later on I discovered that the harder problem was knowing what the problems were and how to use the information. Now, on my desk, there appears to be another source of all knowledge. The formulae from *The Handbook* and much more are on my computer in a superb program called *Mathematica*. A substantial part of known mathematics is built in. I haven't been so excited about a technical program since I first saw *MacPaint* by Bill Atkinson of Apple Computer, Inc.

How can you use *Mathematica*? Believe it or not this tool is recommended for a range of users from grade school to senior scientists. It does numerical, symbolic, and graphical computation. Almost any finite problem I thought of could be handled by this program. You type in the problem and push Enter. Be sure to use the computer standard notation for arithmetic: +, −, * for multiplying, and / for division. Put comments in (*comment*). Start with

```
2 + 2   (*warm-up*)
4
```

and experiment with some harder arithmetic.

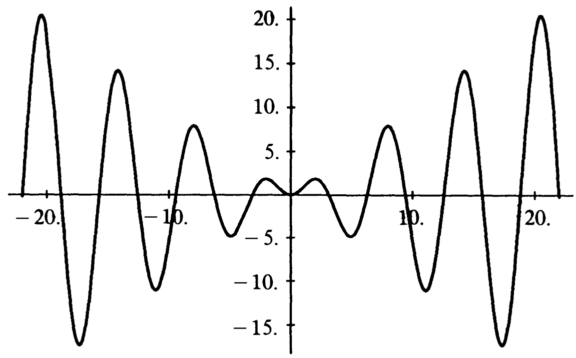
```
123 (4^56)   (*Note that "^" is exponentiation*)
638625251359978379830925104849407180 8
```

I wondered about irrational numbers—pi, in particular. I remember memorizing Pi to 15 or 20 places to impress the girls in my geometry class. Can this program do better? Since *Mathematica* capitalizes its known constants I obtained Pi to 1000 places just by typing in `N[Pi, 1000]`. The N denotes a numerical constant is wanted, and the capital letter in Pi indicates a constant. Given enough memory this program could actually calculate Pi to hundreds of thousands of places. I'll spare you the 1000-place answer. The beauty of this program is that you can use it like a calculator and not worry about writing any computer programs. I requested the square root of 93.956 to 50-digit accuracy which took the program less than one-half of a second to execute. I entered `N[Sqrt[93.956], 50]` and out came `9.693090322....` Factoring is easy. All you have to do to factor $y^{18} - 1$ is type `Factor [y^18 - 1]` and get `(1 + y`$^3$` + y`$^6$`)(1 - y`$^3$` + y`$^6$`)(1 + y + y`$^2$`)(1 - y + y`$^2$`)(1 + y)( - 1 + y)`. I decided to find the prime factors of a sample integer: `FactorInteger[487534432]` returned $(2^5)(7)(11)(17)(103)(113)$ which would

have taken me all day. Note how easy the command language is. It's mostly English.

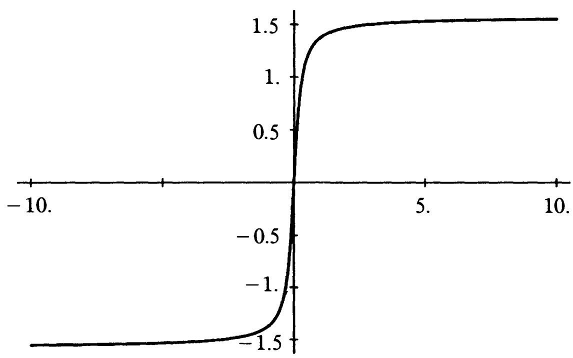
*Mathematica* really shines on its graphics where simple trig functions come to life

```
Plot[x Sin[x], {x, -22, 22}]
```




and you can quickly change the period and amplitude to get the feel of a function.

```
Plot[ArcTan[5x], {x, -10, 10}]
```




My sense of trigonometry in high school was mostly symbolic, not visual, and I remember over half the class getting lost on arctan. This program will be a boon to math teachers. In calculus the possibilities are endless. Just think of the fun you can have plotting conic sections! Circles, ellipses, strange polygons, control over parallel sides,... there is a whole new series of textbooks with accompanying computer programs waiting to be written on this. They will tie together integration and differentiation graphically to make the concepts more understandable. Even the technical math definitions might become clear. Remember the definition of continuity? For every epsilon > 0 there is a delta such that... I don't think I really mastered that until I was in grad school. Epsilon and delta can be visualized as segments or as circle radii, and *Mathematica* can turn them into pictures. Since *Mathematica* can solve most textbook problems in algebra and calculus, teachers

will have to rethink homework assignments. No more grinding out limits. This program is to mathematics what the calculator is to arithmetic. Does anybody bother to learn how to calculate square roots anymore? We might be able to avoid the busy work of calculation and get down to the more important questions. How does it work, and how can we use it? It bothered me when I earned my BS degree by memorizing so many formulae without having a sense that I could solve real world problems. I think I needed a course taught with *Mathematica*.

I explored the program a little more deeply and was able to calculate a Fourier transform, do matrix inverses and determinants, do some partial differentiation, calculate some limits, and deal handily with complex numbers. Don't underestimate the ability to find matrix inverses. This is equivalent to solving complicated math problems and can save someone incredible time.

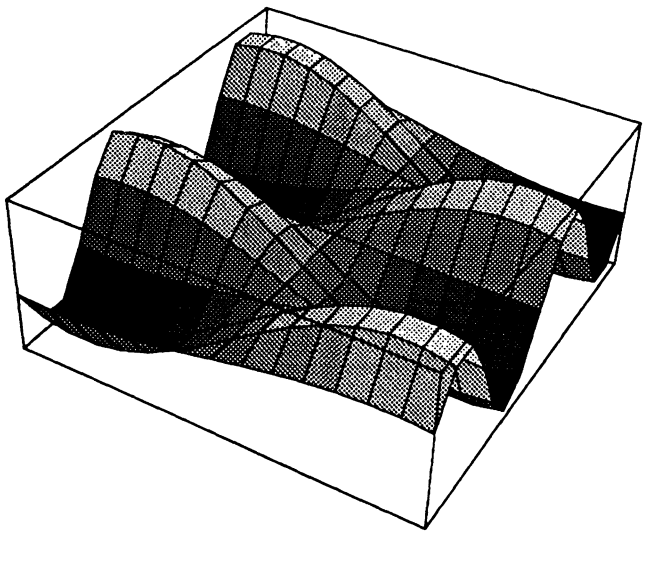
```
  (*Mathematica can work with vectors and matrices,
  which are represented as lists and lists of lists, re-
  spectively. Here is a list of perfect cubes. Table gen-
  erates a list of the values of x^3, with x ranging from
  2 to 10.*)
tbl = Table[x^3, {x, 2, 10}]
{8, 27, 64, 125, 216, 343, 512, 729, 1000}
(*Arithmetic operations are applied to each element on a
list.*) tbl - 1
{7, 26, 63, 124, 215, 342, 511, 728, 999}
(*Select the third number on the list. Double brackets are
used for indexing.*) tbl[[3]] 64
(*Create a matrix*) powers = Table[x^y, {x, 1, 4}, {y, 1, 4}]
{{1, 1, 1, 1}, {2, 4, 8, 16}, {3, 9, 27, 81}, {4, 16, 64,
256}}
(*Display it prettily*) TableForm[powers]

                    1     1     1     1

                    2     4     8     16

                    3     9     27    81

                    4     16    64    256

(*Invert it*) Inverse[powers]

              4      1           13   19     7   11
     {{4, -3, -, -(-)}, {-(--), --, -(-), --},
              3      4           3    4      3   24

       3      7     1         1   1     1   1
      {-, -2, -, -(-)}, {-(-), -, -(-), --}}
       2      6     4         6   4     6   24

(*This is wonderful! I find calculating matrix inverses to
be a drag.*)
```

There is an animation capability that looks clever, and I watched two wave forms move like the ocean. The number of calculations must be immense and the only slowdowns came as I heard the computer make its disk accesses because of the limited memory available in RAM.

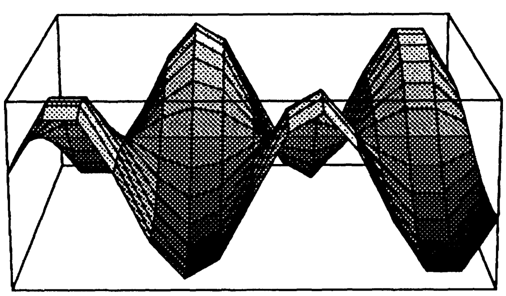
My favorite use of *Mathematica* is its 3D plots which can be done in glorious colors of the user's choice and from any viewpoint. In addition one can obtain contour plots, point plots and much more. Too bad you only see this review in black and white. In color it is magnificent. The 2- and 3-dimensional graphics are generated in PostScript form. This means that they are not held as individual points but as vectors. You can zoom in or out on any of them, stretch and distort them to your heart's content. The graphics can be mixed with text to create well-integrated documents. At a talk I papered the walls with large posters of 3D graphics. With the free PostScript expansion I made the graph of the Gamma function print out on a 6-foot by 8-foot dramatic poster.

Here are a few examples of 3D plots:

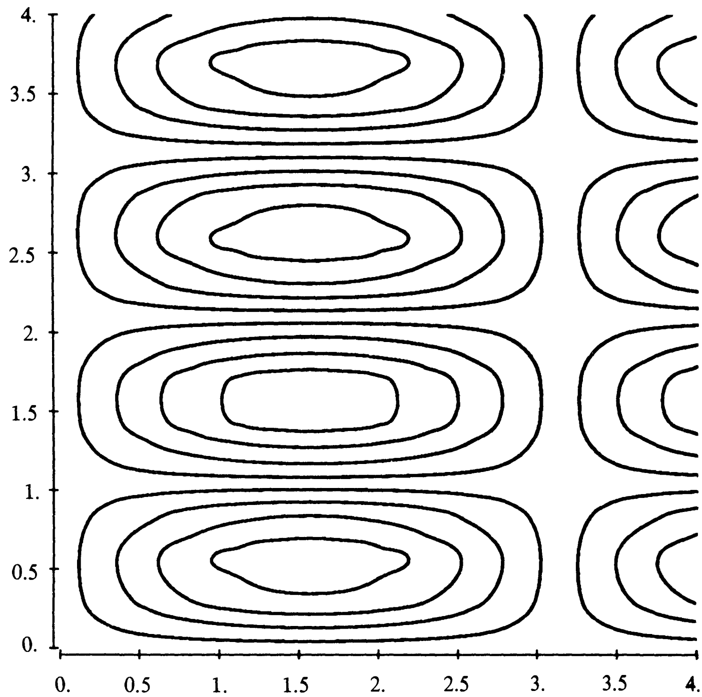
```
Plot3D[Sin[x] Sin[3y], {x, -2, 2}, {y, -2, 2}]
```



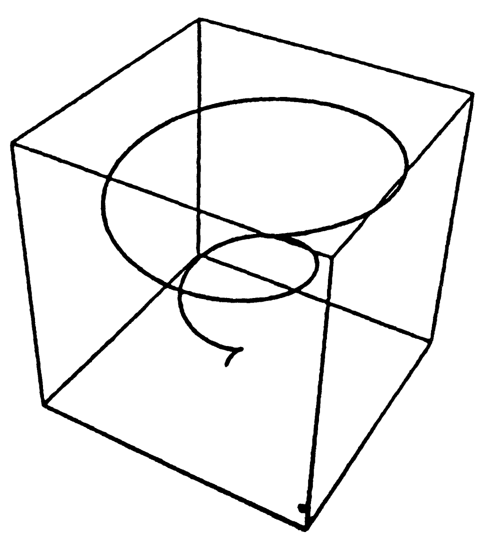
```
Show[%, ViewPoint->{4, 0, 1}]
```



```
ContourPlot[Sin[x] Sin[3y], {x, 0, 4}, {y, 0, 4}]
```




```
SpaceCurve[{u Sin[u], u Cos[u], u},
   {u, 0, 15, 0.15}, BoxRatios- > {1, 1, 1}]
```



research is in fractal computations I look forward to this program running on the new Mac IIx with its even faster chips.

**Criticism.** *Mathematica* is very fast, though you must be patient if you give it problems in the dreamy realms. It took 1.55 seconds to calculate Pi to 1000 places, but took over a minute when I asked for 2000 place accuracy. Asking the program to expand $(1 + x)$ to the power of 10 to the 100th is not fair. This enormous exponent (which is named a "googol") is more than the total number of particles in the universe. The program needs a minimum of 2 megabytes to run the kernel, and I had difficulties loading the integral library. To generate the gamma function, for example, you need 4 MB of memory. If you have over 2.5 megs the program really starts to cook. If *Mathematica* runs out of memory during a calculation it typically crashes. My manual said I could terminate a calculation with the control key, but this didn't work for me.

**Cost.** *Mathematica* is new and costly to retail purchasers. The price is \$495 for the standard version, \$795 for the Mac II version. Since it will be standard on the NeXT machine, Steve Jobs obviously feels it is worthwhile. I understand it will be available to the educational market at substantial discounts. According to Jim Finn, technical editor of MacUser magazine, this is one of the most complex pieces of software ever written. The size of the program is 1.8MB of C code, (it came on 5 floppy disks) so you need a hard disk.

**Where to get it.** Write Wolfram Research Inc., P.O. Box 6059, Champaign, IL 61821 or call 1-800-441-MATH.

*Dynamical Systems.* Ed: V. I. Arnold. Encyclopedia of Mathematical Sciences, V. III. Springer-Verlag, Berlin, 1988. xiv + 291 pp.

John Hornstein
*Naval Research Laboratory, Washington, DC 20375*

Here is an example of how differential geometry, differential and algebraic topology, and Newton's laws make music together. The example is from the book under review.

Imagine a mechanical system whose possible configurations are the points of a smooth 2-dimensional orientable surface. For example, a point on a torus could represent a configuration of a planar double pendulum, or the momentary location of a single mass constrained to move on the torus. In general, the surface is called the configuration manifold.

As Jacobi discovered long ago [**J1**, **J2**], if mechanical energy is conserved during the motion, the configuration point will trace out a geodesic. The line element is given by $(ds)^2 = 2(E - U)\Sigma A_{ij}\, dq_i\, dq_j$. The $q_i$ are the local coordinates, the $\mathring{q}_i$ are the corresponding velocities, $T = (1/2)\Sigma A_{ij}\mathring{q}_i\mathring{q}_j$ is the kinetic energy, $E$ is the total energy, and $U$ is the potential energy. Both $A_{ij}$ and $U$ can depend on location on the configuration manifold. Notice a peculiar feature: the line element depends on the value of the energy. To have a fixed metric for the surface, we must restrict attention to systems all having the same energy. If $E$ is large enough, the factors $T$ and $E - U$ will be positive everywhere. The line element will then define a Riemannian metric over the entire configuration manifold. Suppose this to be the

research is in fractal computations I look forward to this program running on the new Mac IIx with its even faster chips.

**Criticism.** *Mathematica* is very fast, though you must be patient if you give it problems in the dreamy realms. It took 1.55 seconds to calculate Pi to 1000 places, but took over a minute when I asked for 2000 place accuracy. Asking the program to expand $(1 + x)$ to the power of 10 to the 100th is not fair. This enormous exponent (which is named a "googol") is more than the total number of particles in the universe. The program needs a minimum of 2 megabytes to run the kernel, and I had difficulties loading the integral library. To generate the gamma function, for example, you need 4 MB of memory. If you have over 2.5 megs the program really starts to cook. If *Mathematica* runs out of memory during a calculation it typically crashes. My manual said I could terminate a calculation with the control key, but this didn't work for me.

**Cost.** *Mathematica* is new and costly to retail purchasers. The price is $495 for the standard version, $795 for the Mac II version. Since it will be standard on the NeXT machine, Steve Jobs obviously feels it is worthwhile. I understand it will be available to the educational market at substantial discounts. According to Jim Finn, technical editor of MacUser magazine, this is one of the most complex pieces of software ever written. The size of the program is 1.8MB of C code, (it came on 5 floppy disks) so you need a hard disk.

**Where to get it.** Write Wolfram Research Inc., P.O. Box 6059, Champaign, IL 61821 or call 1-800-441-MATH.

*Dynamical Systems.* Ed: V. I. Arnold. Encyclopedia of Mathematical Sciences, V. III. Springer-Verlag, Berlin, 1988. xiv + 291 pp.

John Hornstein
*Naval Research Laboratory, Washington, DC 20375*

Here is an example of how differential geometry, differential and algebraic topology, and Newton's laws make music together. The example is from the book under review.

Imagine a mechanical system whose possible configurations are the points of a smooth 2-dimensional orientable surface. For example, a point on a torus could represent a configuration of a planar double pendulum, or the momentary location of a single mass constrained to move on the torus. In general, the surface is called the configuration manifold.

As Jacobi discovered long ago [**J1**, **J2**], if mechanical energy is conserved during the motion, the configuration point will trace out a geodesic. The line element is given by $(ds)^2 = 2(E - U)\Sigma A_{ij}\, dq_i\, dq_j$. The $q_i$ are the local coordinates, the $\mathring{q}_i$ are the corresponding velocities, $T = (1/2)\Sigma A_{ij}\mathring{q}_i\mathring{q}_j$ is the kinetic energy, $E$ is the total energy, and $U$ is the potential energy. Both $A_{ij}$ and $U$ can depend on location on the configuration manifold. Notice a peculiar feature: the line element depends on the value of the energy. To have a fixed metric for the surface, we must restrict attention to systems all having the same energy. If $E$ is large enough, the factors $T$ and $E - U$ will be positive everywhere. The line element will then define a Riemannian metric over the entire configuration manifold. Suppose this to be the

case, purely for simplicity. If there are no forces other than those confining the particle to the surface, then $U$ is constant; if $A_{ij}$ is also constant then the metric reduces (apart from a constant scale factor) to that induced from an embedding Euclidean space. The geodesic character of the trajectories for this case was known to Euler in 1736. When $U$ is not constant the metric differs from that induced by an embedding Euclidean space. Force then affects the motion by altering the geometry, somewhat as in general relativity.

Now suppose the surface to be connected, compact and without boundary. The surface will be diffeomorphic to a sphere with $h$ handles, and its Euler-Poincaré characteristic will be $\chi = 2 - 2h$. The Euler-Poincaré characteristic is related to the surface average of the Gaussian curvature: $\int K\,dA = 2\pi\chi$. (This is a special case of the Gauss-Bonnet theorem. $K$ is the local value of the Gaussian curvature product, so the left-hand side is the surface area times the average Gaussian curvature.) If $h$ is 0, then $\chi$ is positive, if $h$ is 1 then $\chi$ vanishes, and if $h$ exceeds 1 then $\chi$ is negative; and similarly for the average curvature product.

Where the curvature is positive the surface looks locally like a small piece of an ellipsoid; geodesics diverging from a point tend to converge again, like meridians diverging from the Earth's North pole and converging at the South pole. Where the curvature product is negative the surface looks locally like a saddle-point; diverging geodesics diverge exponentially, at least locally. The Gauss-Bonnet theorem thus shows that, on the average, focusing predominates on a diffeomorph of a sphere, focusing and defocusing balance on a torus, and defocusing predominates if the surface has two or more handles.

The different handles of a multihandled surface offer a choice of paths. Moreover, multihandled surfaces typically have a patch of negative curvature near the attachments of handles. This conjunction of choice and sensitivity to initial conditions suggests that motion on a multihandled surface will in some sense be wilder than motion on a sphere or a torus.

A detailed analysis [**K1**, **K2**] based on homology bears this out: Some force laws on diffeomorphs of a sphere or a torus yield integrable mechanical systems, but no force law can yield an integrable system on a multihandled surface. The possibility of integrability is constrained by the topology of the configuration manifold.

Over the past two decades the theory of dynamical systems has flowered. Deterministic chaos and unexpected regularity (solitons, and regularities within chaos) have provided new paradigms for mathematical studies and for applications. As suggested by the previous example, differential and algebraic topology and differential geometry have contributed strongly to this flowering. Point-set topology and measure theory also contributed in a major way, and algebraic geometry has played a role in studies of unexpected integrability.

*Dynamical Systems III* is one volume of the 5-volume *Encyclopedia of Mathematical Sciences*. The Encyclopedia surveys mathematical advances in the theory of dynamical systems. Other volumes in the series cover ordinary differential equations, smooth dynamical systems, ergodic theory, statistical mechanics and kinetic equations, symplectic geometry, geometric quantization, integrable systems, and bifurcations and catastrophes. (Many of these topics have applications outside of physics, despite their names. For example, statistical mechanics is useful in studies of optimization and automata; kinetic equations can be applied to stochastic processes in general.) *Dynamical Systems III* consists of a single long article: "Mathematical Aspects of Classical and Celestial Mechanics," by V. I. Arnold,

V. V. Kozlov, and A. I. Neishtadt. The book will therefore be called AKN in the remainder of this review.

Arnold, the editor and first author, has contributed strongly to the flowering of the theory of dynamical systems. A student of Kolmogorov, he played a crucial role in the development of the celebrated Kolmogorov, Arnold, Moser (KAM) theory, which proved that fossil traces of integrability remain in non-integrable Hamiltonian systems. He discovered an important general feature (Arnold diffusion) of deterministic chaos in Hamiltonian systems having more than two degrees of freedom. He is known for the "Arnold tongues" in mode-locked systems. Arnold is also responsible for some of the main results in the theory of stable singularities of differentiable mappings. (Catastrophe theory is part of the latter theory.) He has also done major work unrelated to dynamical systems, notably joint work with Kolmogorov on "superpositions," which disproved the conjecture underlying Hilbert's 13th problem. Arnold has written widely-admired expositions on mechanics [**A1**] and ordinary differential equations [**A2**, **A3**], coauthored a classic monograph [**AA**] on ergodic theory in classical mechanics, and is single or joint author of several books on the stable singularities of differentiable mappings.

As might be expected from this pedigree, AKN's emphases and many of its treatments are unique. AKN is especially strong on integrable versus nonintegrable systems, and on the method of averaging and other sophisticated improvements to classical perturbation theory. There is a particularly nice survey of KAM theory. AKN points out a remarkable feature of KAM theory: the results are valid over infinite time, unlike results obtainable from the method of averaging. Other themes are integrable versus nonintegrable constraints; the benefits of symplectic geometry in analyzing Hamiltonian systems; and the proper exploitation of symmetries.

One novel feature of the book is its discussion of "vakonomic" mechanics, a generalization of ordinary Hamiltonian mechanics to handle non-integrable constraints. Vakonomic mechanics was developed by Kozlov, the second author of AKN, and AKN provides the first treatment of this material in book form. I must confess to initial skepticism about vakonomic mechanics. The theory embeds prior theories of systems having non-integrable constraints into a continuum of theories, and asserts that the choice of the proper equation of motion must be made empirically for each mechanical system. The latter assertion especially raised my hackles, because it seemed to contradict the reason for having a theory. The last section in Chapter 1 of AKN and comments on page 96 of [**A1**] changed my mind, convincing me that vakonomic mechanics is an important contribution. Since the essential issue has analogs in mathematical modelling in general, it is worth taking a moment to explain it. The constraints pertinent to the present context occur in simplified, approximate models of systems. Examples: a nearly rigid ball rolling on a wooden floor is idealized as a rigid ball on a rigid floor, subject to rolling friction; a skate or ski digging into ice or snow is idealized as a rigid linear object able to move only along its length, over rigid ground. The constrained system arises as the limit of a more detailed model when one or more parameters (masses, stiffnesses, viscosities) become either overwhelmingly large or ignorably small compared to other quantities in the problem. In many cases the final model depends on the order in which these limits are taken. There are therefore a variety of mathematical models for the simplified system, and choosing the correct one depends either on an analysis of an underlying less-simplified model or on empirical observation. In either case, we are forced to temporarily venture outside of the universe of discourse

of the simplified model. (Most physicists are comfortable with a shifting universe of discourse; most mathematicians are not. This appears to be one of the biggest temperamental differences between the two.)

As an encyclopedia article, AKN does not seek to serve as a textbook, nor to replace the original articles whose results it describes. The book's goal is to provide an overview, pointing out highlights and unsolved problems, and putting individual results into a coherent context. It is full of historical nuggets, many of them surprising. Many results are stated without proof; proofs are given only when they are short and elegant. Although there are no formal prerequisites, much of the material is advanced and even recent. (Many of the references are to works published in the present decade.) The examples are especially helpful; if a particular topic seems difficult, a later example frequently tames it. The writing is refreshingly direct, never degenerating into a vocabulary lesson for its own sake.

The book accomplishes the goals it has set for itself. While it is not an introduction to the field, it is an excellent overview. Readers who become intrigued by the subject will want to read Arnold's text on mechanics [**A1**] and his two books on ordinary differential equations [**A2**, **A3**] before their second pass through AKN. (Both books on ordinary differential equations emphasize dynamical systems.)

Of course, classical mechanics is now only one corner of the theory of dynamical systems. While Hamiltonian systems have unexpectedly wide scope (for example, integrable partial differential equations, and control theory), they have features which are not typical of dynamical systems in general. For example, dissipative dynamical systems can have attractors, but nondissipative systems cannot. (In particular, deterministic chaos in Hamiltonian systems occurs without benefit of a strange attractor, contrary to a popular misconception.) A reader seeking an introduction to the broad field of dynamical systems should consult Guckenheimer and Holmes [**GH**] and Lichtenberg and Lieberman [**LL**] on deterministic chaos, and Drazin [**D**] for a concise overview of solitons.

## REFERENCES

A1. V. I. Arnold, Mathematical Methods of Classical Mechanics (2nd corrected printing), Springer-Verlag, New York, 1980.

A2. ______, Geometrical Methods in the Theory of Ordinary Differential Equations, Springer-Verlag, New York, 1983.

A3. ______, Ordinary Differential Equations, MIT Press, Cambridge, Mass., 1973.

AA. V. I. Arnold and A. Avez, Ergodic Problems of Classical Mechanics, Addison-Wesley, Reading, Mass., 1968 (reprinted 1989).

D. P. G. Drazin, Solitons, Cambridge University Press, Cambridge, 1983.

GH. J. Guckenheimer and P. Holmes, Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields, Springer-Verlag, New York, 1983.

J1. C. G. J. Jacobi, Note sur l'intégration des equations differentielles de la dynamique, *Comptes rendus de l'Academie des sciences de Paris*, V (1837) 61–67. (Reprinted as pp. 129–136 in vol. IV of C. G. Jacobi's Gesammelte Werke, G. Riemer, 1881–91; The Gesammelte Werke has been reprinted by Chelsea Pub. Co., New York.)

J2. ______, Vorlesungen Über Dynamik (second rev. ed.), 1866 and 1884. (Reprinted by Chelsea Pub. Co., New York, 1969.)

K1. V. V. Kozlov, Topological obstructions to the integrability of natural mechanical systems, *Sov. Math. Doklady*, 20(6), 1413–1415 (1979) (English translation of *Dokl. Akad. Nauk SSSR*, 249(6) (1979).

K2. ______, Integrability and non-integrability in Hamiltonian systems, *Russian Mathematical Surveys*, 38(1) 1–76 (1983). (Translation of *Uspekhi Mat. Nauk*, 38:1 3–67 (1983).

LL. A. J. Lichtenberg and M. A. Lieberman, Regular and Stochastic Motion, Springer-Verlag, New York, 1983.

# TELEGRAPHIC REVIEWS

*Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.*

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

| | | | | | |
|---|---|---|---|---|---|
| T: | Textbook | P: | Professional Reading | 1–4: | Semesters |
| C: | Computer Software | L: | Undergraduate Library | **: | Special Emphasis |
| S: | Supplementary Reading | 13: | Grade Level | ??: | Questionable |

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

**Mathematics Appreciation, S, L.** *Fascinating World of Mathematical Sciences, Volume I-III.* J.N. Kapur. Mathematical Sciences Trust Society (New Friends Colony, New Delhi), 1989, RS 95 each. *Volume-I: Nature of Mathematics,* xiv + 264 pp; *Volume-II: Applications of Mathematics—1,* xv + 272 pp; *Volume-III: Applications of Mathematics—2,* xvi + 256 pp. First three of eight volumes reprinting Kapur's numerous papers on the nature of mathematics. Papers in these volumes include 26 reflections on mathematics and 50 illustrations of applications (modelling, life sciences, social sciences, management). Level of exposition ranges from nontechnical after dinner talks to detailed presentations of specific models. Subsequent volumes will feature mathematics education, biography, history, and reviews. LAS

**Precalculus, T(13: 1).** *College Algebra, Fifth Edtion.* Margaret L. Lial, Charles D. Miller. Scott Foresman, 1989, 570 pp, $25. [ISBN: 0-673-38245-1] This edition includes algebra review in a new Chapter 1; new applications of quadratic equations; rearrangement of chapters to focus on relations and functions; treatment of exponential and logarithmic functions now follows polynomials and uses calculators; and rearrangement of the material on matrices and systems. (*First Edition,* TR, October 1973; *Second Edition,* TR, August-September 1978; *Third Edition,* TR, August-September 1981; *Fourth Edition,* TR, January 1986.) JNC

**Precalculus, T(13: 1).** *College Algebra with Applications, Fourth Edition.* Bernard J. Rice, Jerry D. Strange. Brooks/Cole, 1989, xii + 398 pp, $36. [ISBN: 0-534-10206-9] Central theme is still functional concepts. Changes from previous editions (*First Edition,* TR, August-September 1978; *Third Edition,* TR, October 1987) include some reorganization. For example, review of high school skills is now contained in one chapter instead of two. New to this edition is a chapter on vectors and complex numbers. Also, a section on polar coordinates and equations has been added to the chapter on conic sections. GN

**Finite Mathematics, T.** *Finite Mathematics with Calculus: An Applied Approach.* David E. Zitarelli, Raymond F. Coughlin. Saunders College, 1989, xxi + 966 pp, $35. [ISBN: 0-03-011274-5] For business, social, and biological science majors; includes algebra review, finite mathematics (linear mathematics, probability and statistics and applications) and calculus (including several variables). Sections contain case studies and referenced applications; supplements include a computer supplement, computerized test bank, and prepared tests. JNC

**Education, S, P, L*.** *The Teaching and Assessing of Mathematical Problem Solving.* Ed: Randall I. Charles, Edward A. Silver. Res. Agenda for Math. Educ., Volume 3. NCTM and Lawrence Erlbaum Associates, 1988, x + 282 pp, $49.95; $15 (P). [ISBN: 0-8058-0355-6; 0-87353-267-8] Revised versions of papers presented at a working conference held in January 1987 in San Diego, one of whose main purposes was to open new avenues for research in teaching and evaluating mathematical problem solving. Seventeen papers by such authors as Greeno, Resnick, Schoenfeld, and Kilpatrick provide a sampling of perspectives and research methodologies ranging from epistemology and metacognition to assessment and pedagogy. Part of a series designed to support implementation of the new NCTM *Curriculum and Evaluation Standards for School Mathematics.* LAS

**Foundations, P, L.** *Rereading Russell: Essays in Bertrand Russell's Metaphysics and Epistemology.* Ed: C. Wade Savage, C. Anthony Anderson. Minn. Stud. in the Philo. of Sci., V. XII. U of Minnesota Pr, 1989, xii + 320 pp, $29.50. [ISBN: 0-8166-1649-3] Fifteen reflections on Russell's philosophy of science focused on his 1927 book *The Analysis of Matter,* its precursors (including the 1901 *Principia Mathematica*), and its successors (including the 1959 autobiographical *My Philosophical Development*). Begins with three papers on ramification, on the theory of

types, and on Russell's paradox; other papers move into Russell's later work on science, metaphysics, and epistemology. LAS

**Graph Theory, P.** *Graphs and Algorithms.* Ed: R. Bruce Richter. Contemp. Math., V. 89. AMS, 1989, xv + 197 pp, $25 (P). [ISBN: 0-8218-5095-4] Proceedings of a 1987 AMS-IMS-SIAM Joint Summer Research Conference held in Boulder, Colorado convened to foster communication between computer scientists and mathematicians who work in graph theory and related algorithms. Thirteen papers plus a collection of open questions. SB

**Calculus, S(13-14).** *Higher Mathematics.* V.S. Shipachev. Ed: A.N. Tikhonov. MIR (US Distr: Imported Pub), 1988, 512 pp, $15.95. [ISBN: 5-03-000233-2] Single and multiple variable calculus with differential equations through second order linear. Many worked-out examples but no exercise or problem sets. Carefully done, from simple examples to more complicated ones. Would serve as a useful reference for those wishing to supplement an assigned textbook. Few topics in a standard calculus course are missing. The translation is smooth, and the book is very readable. JK

**Real Analysis, T(16-18), L.** *Continuity, Integration and Fourier Theory.* Adriaan C. Zaanen. Universitext. Springer-Verlag, 1989, viii + 251 pp, $39 (P). [ISBN: 0-387-50017-0] Aimed at students who may not have been introduced to Lebesgue integration, but who need it for Fourier series and transforms. "In view of the emphasis in modern mathematics curricula on abstract subjects—it may be useful to have a textbook available (not too elementary and not too specialized) on the subjects—classical but still important today—which are mentioned in the title of this book." AWR

**Real Analysis, T(17-18: 2), L.** *Real Analysis, Third Edition.* H.L. Royden. Macmillan, 1988, xvii + 444 pp, $38. [ISBN: 0-02-404151-3] This is a revision with a purpose. Previous treatment of Baire and Borel measures in locally compact spaces, "seriously flawed" in the author's analysis of his *Second Edition* (TR, October 1968), is here corrected. A new chapter on invariant measures, and improvements in both text and problems seem to insure that libraries will want to keep this venerable text in easy reach of students for another twenty years. AWR

**Differential Equations, T(18: 1, 2), P.** *Differential Equations with Discontinuous Righthand Sides.* A.F. Filippov. Math. & Its Applic. Kluwer Academic, 1988, x + 304 pp, $99. [ISBN: 90-277-2699-X] Many results in the theory of differential equations are shown to be valid for those equations with discontinuous righthand sides. Properties of solutions due to this discontinuity are discussed. GN

**Differential Equations, P.** *Mathematical Aspects of Vortex Dynamics.* Ed: Russel E. Caflisch. SIAM, 1989, 220 pp, $29.50 (P). [ISBN: 0-89871-235-1] Includes nineteen papers presented at the workshop on mathematical aspects of vortex dynamics held in Leesburg, Virginia in April 1988. GN

**Operator Theory, T(18: 2), P.** *Vertex Operator Algebras and the Monster.* Igor Frenkel, James Lepowsky, Arne Meurman. Pure & Appl. Math., V. 134. Academic Pr, 1988, liii + 502 pp, $69.96. [ISBN: 0-12-267065-5] Develops mathematical connection between certain vertex operator algebras and string theory. Introduces affine Kac-Moody algebras and vertex operator algebras, then constructs the Monster and the moonshine module and corresponding vertex operator algebra. Assumes abstract algebra at the graduate level. SB

**Analysis, P.** *Function Classes of Cauchy Continuous Maps.* Eva Lowen-Colebunders. Pure & Appl. Math., V. 123. Marcel Dekker, 1989, xiv + 166 pp, $89.75. [ISBN: 0-8247-7992-4] Discusses algebraic and topological structures of an extension $Y$ of a space $X$ by means of a suitable Cauchy structure on $X$. By viewing the class of real-valued functions on $X$ which are continuously extendable to $Y$ as the class of real-valued Cauchy continuous maps of the related Cauchy space, the author combines the theory of extensions and function classes with the theory of Cauchy spaces and Cauchy continuous maps. GN

**Analysis, P.** *Lecture Notes in Mathematics-1315: Topological Properties of Spaces of Continuous Functions.* Robert A. McCoy, Ibula Ntantu. Springer-Verlag, 1988, iv + 124 pp, $13.10 (P). [ISBN: 0-387-19302-2] For most topologies imposed on $C(X,R)$ (the space of all continuous functions from $X$ into $R$), the topological properties of $X$ and $R$ interact with those of $C(X,R)$. This text studies these interactions, especially the cases of set-open topologies and uniform topologies on $C(X,R)$. Contains exercises at the end of each chapter. GN

**Analysis, P.** *Algebraic Analysis: Papers Dedicated to Professor Mikio Sato on the Occasion of His Sixtieth Birthday, Volume I & II.* Ed: Masaki Kashiwara, Takahiro Kawai. Academic Pr, 1988, $59.95 each. *Volume I,* xxi + 472 pp [ISBN: 0-12-400465-2]; *Volume II,* xxi + 473 pp. [ISBN: 0-12-400466-0] A collection of over sixty contributed papers discussing various results in algebraic analysis. GN

**Differential Geometry, P.** *Lecture Notes in Mathematics-1369: Differential Geometry and Topology.* Ed: Boju Jiang, Chia-Kuei Peng, Zixin Hou. Springer-Verlag, 1989, vi + 366 pp, $34.90 (P). [ISBN: 0-387-51037-0] A collection of twenty-four papers and lecture notes from a Special Year in Geometry and Topology, held in 1986-87 at Nankai, China. The first two long papers on Dupin submanifolds and curvature of tubular hypersurfaces respectively are expository in style. PZ

**Topology, P.** *Minimal Flows and Their Extensions.* Joseph Auslander. Math. Stud., V. 153. North-Holland (US Distr: Elsevier Science), 1988, xi + 265 pp, $86.75. [ISBN: 0-444-70453-1] A flow is a jointly continuous action of a topological group on a compact Hausdorff space. This account is set in the context of the AMS Colloquium book by Gottschalk and Hedlund, *Topological Dynamics,* and overlaps in content with Ellis, *Lectures in Topological Dynamics,*

Benjamin, 1969, and Bronstein, *Extensions of Minimal Transformation Groups*, Sitjthoff and Noordhoff, 1979. Note the price of this typed-for-camera manuscript. AWR

**Operations Research, P.** *Operations Research Models in Flexible Manufacturing Systems*. Ed: F. Archetti, M. Lucertini, P. Serafini. CISM Courses & Lectures, No. 306. Springer-Verlag, 1989, 305 pp, $43.40 (P). [ISBN: 0-387-82099-X] Papers which stay close to actual applications given at a conference involving both academicians and industrial workers. Flexible manufacturing systems are characterised by families of related parts that can be made simultaneously, and by manufacturing processes in which change-over time between operations is small compared to the operating times. Simulation, planning, and scheduling all come in for attention. AWR

**Optimisation, T(18: 1), L.** *Theory of Duality in Mathematical Programming*. Manfred Walk. Springer-Verlag, 1989, 178 pp, $49. [ISBN: 0-387-82057-4] Introduces readers to the qualitative theory of mathematical programming in vector spaces. Presents a general duality theory and applies it to several different mathematical programming problems. Prerequisites include basic analysis and linear algebra. No exercises. SM

**Control Theory, P.** *Lecture Notes in Control and Information Sciences-117: Stochastic Optimal Control Theory with Application in Self-Tuning Control*. K.J. Hunt. Springer-Verlag, 1989, x + 308 pp, $44 (P). [ISBN: 0-387-50532-6] Classic linear-quadratic-gaussian (LQG) control theory is here modified to use a polynomial equation approach, obviating the requirement that noise sources be Gaussian distributed, hence the name Stochastic Optimal Control. The book uses this approach together with self-tuning as the adaptive control process. Account is given of application of these techniques to control of steam pressure in the Hunterston, Scotland power station simulator. AWR

**Computational Statistics, T(16-17: 1), P.** *Statistical Modelling in GLIM*. Murray Aitkin, *et al.* Oxford Stat. Sci. Ser., V. 4. Clarendon Pr, 1989, xi + 374 pp, $35 (P); $75. [ISBN: 0-19-852203-7; 0-19-852204-5] Begins with an introduction to GLIM3, a powerful interactive statistical modelling package, and follows with a theoretical chapter on statistical modelling and inference. Then shows how to use GLIM3 to analyze normal regression and analysis of variance models, and models with binomial, multinomial, and Poisson response data. Concludes with a detailed treatment of survival data. RSK

**Statistics, T(14-15: 1).** *Applied Nonparametric Statistical Methods*. Peter Sprent. Chapman & Hall, 1989, x + 259 pp, $32.50 (P). [ISBN: 0-412-30610-7] Practical introduction with a minimum of theory. Much of the presentation is done through examples, using the format: problem, formulation and assumptions, procedure, conclusion, comments. Also includes illustrations from various fields of the type of problem to which the methods could be applied. Assumes some knowledge of elementary statistics. RSK

**Statistics, T(15-16).** *Statistical Analysis: An Interdisciplinary Introduction to Univariate & Multivariate Methods*. Sam Kash Kachigan. Radius Pr, 1986, xviii + 589 pp, $35.95. [ISBN: 0-942154-99-1] A good first course book that introduces multivariate techniques commonly used in many application fields as well as standard univariate techniques. Could be used as a first course in graduate statistics also. Excellent illustrations and reasonable exercises drawn from "behavioral, biological, environmental, and monetary sciences." MS

**Statistics, P*.** *Bayesian Statistics 3*. Ed: J.M. Bernardo, *et al.* Clarendon Pr, 1988, xi + 805 pp, $125. [ISBN: 0-19-852220-7] Proceedings of the Third Valencia International Meeting on Bayesian Statistics held in Spain, in June 1987. Contains all 31 invited papers, with associated discussion, together with a selection of 33 contributed papers. RSK

**Statistics, T(18: 2).** *Characterizations and Analysis of Block Designs*. A.K. Nigam, P.D. Puri, V.K. Gupta. Wiley, 1988, viii + 176 pp, $24.95. [ISBN:0-470-21051-6] Intended for students and researchers interested in incomplete block designs. Mathematics prerequisites are minimized but a certain level of maturity is required to read the text comfortably. Many new techniques are included with central chapters on Kronecker product designs and block designs with factorial structures. No exercises or applications. Poorly bound. MS

**Statistics, T(18: 2), P.** *Robustness of Statistical Tests*. Takeaki Kariya, Bimal K. Sinha. Stat. Modeling & Decision Sci. Academic Pr, 1989, xvi + 189 pp, $44.95. [ISBN: 0-12-398230-8] Primarily concerned with multivariate problems (e.g., GMANOVA and covariance structures). Robustness here means that an optimality property of a test under a normal distribution holds exactly under a non-normal distribution, primarily orthogonally invariant and elliptically symmetrical distributions. Does not use or refer to "breakdown" or "influence curve" side of robustness literature. Considers robustness of null and non-null distributions, and robustness of optimality. TH

**Statistics, P.** *Lecture Notes in Statistics-53: Relations, Bounds and Approximations for Order Statistics*. Barry C. Arnold, N. Balakrishnan. Springer-Verlag, 1989, ix + 173 pp, $20.60 (P). [ISBN: 0-387-96975-6] Recurrence relations, bounds on expectations ("1001 ways to use the Schwarz inequality"), and approximations to moments of order statistics; samples with a single outlier; record values. Extensive references. TH

**Statistics, T(15: 2), L.** *Applied Regression Analysis in Econometrics*. Howard E. Doran. Statistics: Textbooks & Mono., V. 102. Marcel Dekker, 1989, viii + 372 pp, $89.75. [ISBN: 0-8247-8049-3] Requires little statistical background; covers simple and multiple regression, interpretation, hypothesis testing, dummy variables, analysis of residuals, dynamic models and distributed lags, modelling, random $X$ variables, maximum likelihood estimates,

probit models. Few plots and diagrams. Uses Minitab, also Shazam. TH

**Languages, T(16-18: 1).** *Programming Languages: Concepts and Constructs.* Ravi Sethi. Addison-Wesley, 1989, xii + 478 pp. [ISBN: 0-201-10365-6] Starts with role of structure and the basic elements common to all languages. The core of the text treats assignments, activations, encapsulation, inheritance, functional logic, and concurrent programming. Each core topic uses two languages for illustrations (C, Modula-2, Scheme, Prolog, Ada). The last three chapters treat the descriptions of languages. Chapter exercises and bibliographic notes. Bibliography, index. RJA

**Languages, T*(15-16: 1).** *Concepts of Programming Languages.* Robert W. Sebesta. Ser. in Computer Sci. Benjamin/Cummings, 1989, xiv + 497 pp, $33.95. [ISBN: 0-8053-7011-0] Designed for the CS 8 course described in the 1978 ACM Curriculum Recommendations. Devotes a chapter to each of several significant language constructs and examines design choices in several languages. Also includes chapters on functional programming, logic programming, and object-oriented programming. AO

**Theory of Computation, T*(14-16: 1), S, L.** *Theory of Computation: Formal Languages, Automata, and Complexity.* J. Glenn Brookshear. Ser. in Comput. Sci. Benjamin/Cummings, 1989, xii + 322 pp, $37.95. [ISBN: 0-8053-0143-7] Attempts to correlate abstract topics with areas to which they apply. Includes finite automata and regular languages, pushdown automata and context-free languages, Turing machines and phrase-structure languages. Computability is treated in the context of Turing machines, partially recursive functions, and simple programming languages. The third thrust of the text is complexity: computations, algorithms, problems, language recognition, and nondeterministic machines. Appendices, chapter problems, references. RJA

**Theory of Computation, T*(15-17: 1), L*.** *An Introduction to Formal Language Theory.* Robert N. Moll, Michael A. Arbib, A.J. Kfoury. Texts & Mono. in Comput. Sci. Springer-Verlag, 1988, x + 203 pp, $39.90. [ISBN: 0-387-96698-6] Concise, clearly written text for first course in formal languages. Unique feature is two-chapter introduction to formal theory of natural languages. Also covers families of languages and corresponding automata, parsing, computability, fixed point methods. KES

**Theory of Computation, T(16-18: 1, 2), S, P, L.** *PX: A Computational Logic.* Susumu Hayashi, Hiroshi Nakano. Found. of Comput. MIT Pr, 1988, xiv + 200 pp, $30. [ISBN: 0-262-08174-1] Describes a type-free, constructive logic for computation which formalizes a dialect of pure Lisp. Contains formal theory of PX, semantics, program extraction from proofs using PX, as well as implementation and use of PX as a programming language. Appendices, references, index. RJA

**Artificial Intelligence, T(16-18: 1, 2), S, L.** *Formal Methods in Artificial Intelligence.* Allan Ramsay. Tracts. in Theoret. Comput. Sci., V. 6. Cambridge U Pr, 1988, ix + 279 pp, $49.50. [ISBN: 0-521-35236-3] Begins from propositional and predicate calculus; makes a stop at theorem provers for predicate calculus. Next come chapters that extend classical logic: modal logic, temporal reasoning, non-monotonic reasoning, $\lambda$-calculus and Montague semantics. Ends with two alternatives to classical logic: constructive logic and fuzzy logic. Bibliography, index. RJA

**Artificial Intelligence, T*(16-18: 1, 2), S, L.** *Artificial Intelligence and the Design of Expert Systems.* George F. Luger, William A. Stubblefield. Ser. in Artif. Intell. Benjamin/Cummings, 1989, xxv + 660 pp, $36.95. [ISBN: 0-8053-0139-9] The first part presents the breadth and origins of artificial intelligence. Part two deals with representation and search, while the third part comprises a chapter on Prolog and one on Lisp. Part four discusses specific representations for knowledge-based systems. Advanced artificial intelligence programming techniques are given in part five. Chapter exercises and references. Lots of historical information. Interplay between theory and applications. Ends with artificial intelligence as an empirical inquiry. Appendix on Prolog programs and one on Lisp programs. Bibliography, author index, subject index. RJA

**Computer Science, P.** *Foundations of Deductive Databases and Logic Programming.* Ed: Jack Minker. Morgan Kaufmann, 1988, 746 pp, $42.95. [ISBN: 0-934613-40-0] First part treats negation and stratified databases. The second part considers some fundamental issues: use of semantic information to optimize deductive search, development of informative answers to queries, updating integrity constants, deleting data, handling recursive axioms, accessing data and axioms efficiently, implementations. The third part is on unification and logic programming. Concise introduction with references. Lists of authors and referees. Author index; subject index. RJA

**Computer Science, P.** *A Program Generator for Recognition, Parsing and Transduction with Syntactic Patterns.* G.J. van der Steen. CWI Tract, V. 55. Mathematisch Centrum, 1988, 284 pp, Dfl. 42.90 (P). [ISBN: 90-6196-361-3] Focus is on computational aspects of recognition, parsing, and transduction—formalisms suitable for representing various grammars. A formal machine model is described, followed by a concrete implementation and accompanying program generator. Treats complexity of generated programs and various applications. References and index. RJA

**Computer Science, P.** *Constraint Satisfaction in Logic Programming.* Pascal Van Hentenryck. MIT Pr, 1989, xvi + 224 pp, $35. [ISBN: 0-262-08181-4] Presents new approach to solving discrete combinatorial problems. Uses constraints to prune the search space of combinations which cannot appear in a solution. Such search procedures propagate con-

straints as much as possible and assume values of variables (not already assigned) until a solution is found. These notions have been integrated into logic programming and found in systems: CLP, Prolog III, Trilogy. The CHIP project of ECRC is the basis of much of the book. Many applications discussed; references; index. RJA

**Applications, P*.** *Vibration and Coupling of Continuous Systems: Asymptotic Methods.* J. Sanchez Hubert, E. Sanchez Palencia. Springer-Verlag, 1989, xv + 421 pp, $79. [ISBN: 0-387-19384-7] Account of modern methods. From basics to research problems of current interest. Self-contained enough so as to be accessible to engineers and theorists concerned with mechanics as well as specialists in computational methods. Linear problems where elastic structures are coupled to fluids, radiation of immersed bodies, local vibrations, thermal effects, and more. Comments, complements, problems at some chapter ends. Exercises. References to recent publications and to some classics in the literature. JK

**Applications, L**.** *Adventures in Celestial Mechanics: A First Course in the Theory of Orbits.* Victor G. Szebehely. U of Texas Pr, 1989, xiv + 175 pp, $27.50. [ISBN: 0-292-75105-2] Delightful from the Dedication (to the memory of Newton) to the Concluding Remarks 146 pages later. Author's purpose is "to demonstrate the beauty of orbit mechanics and celestial mechanics ..." Emphasis is on the basics. From Earth satellites in circular orbits to a restricted form of the three-body problem. Historical remarks offer insight and background. List of contributors beginning with Aristotle. Roles of Euler, Lagrange, Poincaré, Whittaker, Einstein, and others. The author is a masterful storyteller and his enthusiasm for his subject is catching. Worked examples and problems to solve. Annotated list of major references. In a word, the book is a "winner." JK

**Applications, P.** *Transactions of the Sixth Army Conference on Applied Mathematics and Computing.* US Army Research Office (PO Box 12211, Research Triangle, NC 27709), 1989, xx + 1151 pp, (P). Papers from a June 1988 conference held in Boulder, Colorado. Bound backwards and upside down: move from page 1150 to page 1. LAS

**Applications (Cognitive Science), T(16-18: 1, 2), S, P, L.** *Natural Computation.* Ed: Whitman Richards. MIT Pr, 1988, x + 561 pp, $25 (P). [ISBN: 0-262-68055-6] Contains 36 readings, half on vision. The first section provides an introduction to natural computation, and the last section indicates future directions. In between, the readings are in four groups: information at contours, property tags, sound interpretation, force sensing and control. Each group has an introduction with references. Text concludes with problem sets. Glossary, name and subject indexes. RJA

**Applications (Physics), P.** *Quantum Theories and Geometry.* Ed: M. Cahen, M. Flato. Math. Physics Stud., V. 10. Kluwer Academic, 1988, ix + 191 pp, $69. [ISBN: 90-277-2803-8] Compiled from lectures delivered at the Fondation Les Treilles in March 1987. MU

**Applications (Physics), P*, L*.** *Mathematical Methods of Classical Mechanics, Second Edition.* V.I. Arnold. Transl: K. Vogtmann, A. Weinstein. Grad. Texts in Math., V. 60. Springer-Verlag, 1989, xvi + 508 pp, $49.80. [ISBN: 0-387-96890-3] Based on a year-and-a-half-long course in classical mechanics taught by the author to third- and fourth-year mathematics students at Moscow State University in 1966-1968. Modern approach to Newtonian, Lagrangian, and Hamiltonian formulations of dynamics. Classical mechanics and its relationship to Riemannian geometry, Lie groups, and more. Three new appendices, originally written for inclusion in a German edition, on work by author and co-authors on Poisson structures, elliptic coordinates with applications to integrable systems, and singularities of ray systems. Destined to be a standard in the field (*First Edition,* TR, February 1979). JK

**Applications (Physics), T(18: 1), P.** *Hamiltonian Systems: Chaos and Quantization.* Alfredo M. Ozorio de Almeida. Mono. on Math. Physics. Cambridge U Pr, 1988, ix + 238 pp, $64.50. [ISBN: 0-521-345316] This new English version of the original Portuguese text introduces the theory of classical Hamiltonian systems with main focus on periodic orbits and their neighborhood. MU

**Applications (Physics), P.** *Operator Algebras and Applications.* Ed: David E. Evans, Masamichi Takesaki. Cambridge U Pr, 1988, (P). *Volume 1: Structure Theory; K-Theory, Geometry and Topology.* London Math. Soc. Lect. Note Ser., V. 135, viii + 244 pp [ISBN: 0-521-36843-X]; *Volume 2: Mathematical Physics and Subfactors.* London Math. Soc. Lect. Note Ser., V. 136, viii + 240 pp. [ISBN: 0-521-36844-X] Compiled from papers presented at the UK-US Joint Seminar on Operator Algebras held at the University of Warwick in July 1987. MU

### Reviewers

RJA: Richard J. Allen, St. Olaf; DFA: David F. Appleyard, Carleton; SB: Steve Benson, St. Olaf; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; JD-B: John Dyer-Bennet, Carleton; WE: William Etter, Macalester; SG: Steven Galovich, Carleton; GG: George Gilbert, St. Olaf; RH: Rhonda Hatcher, St. Olaf; TH: Timothy Hesterberg, St. Olaf; PH: Paul Humke, St. Olaf; RJ: Roger Johnson, Carleton; JJ: Jason Jones, St. Olaf; RSK: Richard S. Kleber, St. Olaf; JDEK: Joseph D.E. Konhauser, Macalester; LCL: Loren C. Larson, St. Olaf; SM: Steve McKelvey, St. Olaf; GN: Gail Nelson, Carleton; KO: Kenneth Olstad, St. Olaf; JO: Jeff Ondich, St. Olaf; AO: Arnold Ostebee, St. Olaf; SP: Samuel Patterson, Carleton; MLR: Margaret L. Reese, St. Olaf; MPR: Matthew P. Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; KS: Karen Saxe, St. Olaf; GMS: G. Michael Schneider, Macalester; JS: John Schue, Macalester; JAS: J. Arthur Seebach, Jr., St. Olaf; KES: Kay E. Smith, St. Olaf; LAS: Lynn Arthur Steen, St. Olaf; MSS: Myriam S. Steinback, Macalester; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; MW: Martha Wallace, St. Olaf; PZ: Paul Zorn, St. Olaf.

# Don't Miss These Recent Publications from Springer-Verlag!

## Mathematical Introduction to Linear Programming and Game Theory

By **L. Brickman**

Mathematical elegance is a constant theme in this treatment of linear programming and matrix games. Condensed tableaux, minimal in size and notation, are employed for the simplex algorithm. In the context of these tableaux, the beautiful termination theorem of R.G. Bland is proven more simply than heretofore, and the important duality theorem becomes almost obvious. Examples and extensive discussions throughout the book provide insight into definitions, theorems, and applications. The book is designed for a one-semester undergraduate course.

**1989/130 pp./Hardcover $34.00/ISBN 0-387-96931-4**
***Undergraduate Texts in Mathematics***

## A Course in Modern Geometries

By **J. Cederberg**

The major emphasis of this text is on the geometries developed after Euclid's *Elements* (circa 300 B.C.). In addition it provides an excellent opportunity to explore aspects of the history of mathematics. Also, since algebraic techniques are frequently used, this study demonstrates the interaction of several areas of mathematics and serves to develop geometrical insights into mathematical results which previously appeared to be completely abstract in nature. A course based on this text provides excellent preparation for the standard undergraduate course in abstract algebra.

**1989/app. 232 pp., 133 illus./Hardcover $42.00/ISBN 0-387-96922-5**
***Undergraduate Texts in Mathematics***

## Factorization and Primality Testing

By **D.M. Bressoud**

This self-contained book is about how to factor or prove primality for large numbers; it is also an introduction to the theory of numbers. Intended for upper-level undergraduates in mathematics and computer science the book aims to introduce the current research, give an appreciation for the results that have been produced throughout the history of number theory, and get them using the computer to factor and prove primality. Currently, no introduction to this area is as up-to-date or as accessible to the non-specialist.

**1989/app. 272 pp., 2 illus./Hardcover $45.00/ISBN 0-387-97040-1**
***Undergraduate Texts in Mathematics***

## Polynomials

By **E.J. Barbeau**

Providing a backdrop for later study in calculus, modern algebra, numerical analysis and complex variable theory, this book acts as an extension to the high school curriculum. Its many exercises introduce various techniques and topics in the theory of equations, such as evaluation and factorization of polynomials, solution of equations, interpolation, approximation and congruences. The theory, illustrated through examples rather than given formal treatment, is clearly presented.

**1989/441 pp., 36 illus./Hardcover $59.00/ISBN 0-387-96919-5**
***Problem Books in Mathematics***

## Fractals for the Classroom

By **H.-O. Peitgen, H. Jürgens** and **D. Saupe**
Advisory Board: **E. Maletsky, T. Perciante**, and **L. Yunker**

*Springer-Verlag*, in cooperation with the *National Council of Teachers of Mathematics*, presents this valuable teaching aid for classroom use as well as for independent study. **Fractals for the Classroom** is written especially for teachers and is intended for the high school and college level. It is based on several lectures and lecture series given to various communities of teachers and students. Two special supplementary volumes will soon be available, one which will contain strategic lessons on fractals, and another with strategic computer experiments on fractals.

**1990/app. 160 pp., many illus./Hardcover $29.00 (tent.)**
**ISBN 0-387-97041-X**

---

**Order Today!**
Call Toll-Free: 1-800-SPRINGER (In NJ call 201-348-4033).
Or send FAX: 212-473-6272
For mail orders please send payment plus $2.50 for postage and handling to: **Springer-Verlag New York, Inc.**, Attn.: S. Klamkin-Dept. S337, 175 Fifth Avenue, New York, NY 10010. We accept Visa, MC, and Amex charges (with signature and exp. date noted) as well as personal checks and money orders. NY, NJ, and CA residents please add state sales tax.

---

**Springer-Verlag**
New York Berlin Heidelberg Vienna
London Paris Tokyo Hong Kong

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, DC 20036

# INDEX TO VOLUME 96, 1989
# THE AMERICAN MATHEMATICAL MONTHLY

## TITLE INDEX


An Alternate Proof of the Continuity of the Roots of a Polynomial, FELIPE CUCKER and ANTONIO G. CORBALAN, 342
Another Example of an Exotic Function, VICTOR PAMBUCCIAN, 913
Arakelian's Approximation Theorem, JEAN-PIERRE ROSAY and WALTER RUDIN, 432
The Butterfly Curve, T. H. FAY, 442
Can a Graph be Both Continuous and Discontinuous?, HUGH THURSTON, 814
A Cantor Set of Nonconvergence, DAVID R. ARTERBURN and WILLIAM D. STONE, 604
The Cauchy-Schwarz Inequality: A Geometric Proof, JAMES W. CANNON, 630
A Characterization of a Class of Composition Operators, R. E. LEWKOWICZ, 725
Chebyshev's Inequality and Natural Density, CURTIS N. COOPER and ROBERT E. KENNEDY, 118
Comparing the Spectral Radii of Two Nonnegative Matrices, R. B. BAPAT, 137
Conway's RATS and Other Reversals, RICHARD K. GUY, 425
Counting the Rationals, YORAM SAGHER, 823
Coupled Linear Differential Equations with Real Coefficients, MOGENS E. LARSEN and BJARNE S. JENSEN, 729
Covering Curves by Translates of a Convex Set, K. BEZDEK and R. CONNELLY, 789
Direct Sum of J-Rings and Zero Rings, JIANG LUH and STEVE LIGH, 40
Disks, Balls, and Walls: Analysis of a Combinatorial Game, R. ANDERSON, L. LOVÁSZ, P. SHOR, J. SPENCER, E. TARDOS, and S. WINOGRAD, 481
The Effect of Prior Calculus Experience on 'Introductory' College Calculus, MARTHA B. BURTON, 350
An Eigenvalue Characterization of the Correlation Coefficient, SEYMOUR KASS, 910
An Elementary Test for the Galois Group of a Quartic Polynomial, LUISE-CHARLOTTE KAPPE and BETTE WARREN, 133
An Elementary Treatment of the Radon-Nikodym Derivative, RICHARD C. BRADLEY, 437
Equations in Division Rings—A Survey, J. LAWRENCE AND G. SIMONS, 220
Estimating the Diffusion of Stock Prices with the HP-28S, YVES NIEVERGELT, 636
A Euclidean Model for Euclidean Geometry, ADOLF MADER, 43
Factor Rings of Integers, STEVE JOHNSON, 521
Fair Dice, JOSEPH B. KELLER and PERSI DIACONIS, 337
Fifty Years of Putnam Trivia, JOSEPH A. GALLIAN, 711
Fixed Points of the Twisted Cyclic Shift Operator, LARRY W. CUSICK and PETER TANNENBAUM, 713
A (Fresh)man Treatment of Determinants, KENNETH P. BOGART, 915
From Calculus to Number Theory, JAMES DUEMMEL, 140
From Experimentation to Proof, HERVÉ LEHNING, 631
Fubini's Theorem for Null Sets, ERIC K. VAN DOUWEN, 718
Generalizing the Formula for Areas of Polygons to Moments, S. F. BOCKMAN, 131
A Geometrically Inspired Proof of the Singular Value Composition, S. J. BLANK, NISHAN KRIKORIAN, and DAVID SPRING, 238
Geometry of Continued Fractions, M. C. IRWIN, 696
How Not to Prove Fermat's Last Theorem, JOHN McCLEARY, 410
Irrationals and the Fundamental Theorem of Arithmetic, DAVID J. SPROWS, 732
The Isoperimetric Inequality and Rational Approximation, D. KHAVINSON and T. W. GAMELIN, 18
Kathy O'Hara's Constructive Proof of the Unimodality of the Gaussian Polynomials, DORON ZEILBERGER, 590
Material Implication Revisited, JOSEPH S. FULDA, 247
Merlin's Magic Square Revisited, DANIEL L. STOCK, 608
Minimal Periods of Discrete and Smooth Orbits, STAVROS BUSENBERG, DAVID FISHER, and MARIO MARTELLI, 5
The Missing Boundary of the Blaschke Diagram, J.R. SANGWINE-YAGER, 233
Monotone Multiplicative Functions, JOEL COHEN, 512
A Motivated Proof of the Rogers-Ramanujan Identities, GEORGE E. ANDREWS and R. J. BAXTER, 401
Mountain Climbing, Ladder Moving, and the Ring-Width of a Polygon, JACOB E. GOODMAN, JÁNOS PACH, and CHEE K. YAP, 494
A Natural Interpretation of an Artificial Function, HANSKLAUS RUMMLER, 520
The New Mersenne Conjecture, J. F. SELFRIDGE, P. T. BATEMAN, and S. S. WAGSTAFF, JR., 125
Nonstandard Continuity and Uniform Convergence, CHRISTOPHER L. THOMPSON, 443
The Norm of a Linear Function, I. J. MADDOX, 434

Normalized Symmetric Functions, Newton's Inequalities, and a New Set of Stronger Inequalities, SHMUEL ROSSET, 815
A Note on the Row-Reduction Algorithm, CHIH-HAN SAH, 143
A Note on Taylor's Theorem, JOSE A. FACENDA AQUIRRE, 244
A Note on Venn Diagrams, LEWIS PAKULA, 38
The Number of Words of Length $n$ in a Graph Monoid, DAVID C. FISHER, 610
On $(\mathbf{a} \times \mathbf{b}) \times \mathbf{c}$, RONALD SHAW AND FRED J. YEADON, 623
On a Conjecture of R. J. Simpson About Exact Covering Congruences, DORON ZEILBERGER, 243
On the Differentiation Formula for sin, DONALD HARTIG, 252
On Groups in Which Every Element Has Finite Order, NARAIN GUPTA, 297
On Markov Processes in Elementary Mathematics Courses, JOHN T. BALDWIN, 147
On the Number of Furthest Neighbour Pairs in a Point Set, HERBERT EDELSBRUNNER and STEVEN SKIENA, 614
On Open Maps, JOHN CROWE and DOMINICK SAMPERI, 242
On the Proof of the Radon-Nikodym Theorem, ALBERT WILANSKY, 441
On the Square Roots of Infinite Matrices, LILLIAN E. PETERS HUPERT and ANNE LEGGETT, 34
On the Use of Iteration Methods for Approximating the Natural Logarithm, JAMES F. EPPERSON, 831
On the Vandermonde Matrix, JOSEPH J. RUSHANAN, 921
Optimal Inscribed Polygons in Convex Curves, JOHN S. LEW and DONALD A. QUARLES, JR., 886
An Optimization Problem, RICHARD BASSEIN, 721
Orthogonal Bases of $\mathbb{R}^3$ with Integer Coordinates and Integer Lengths, HANS LIEBECK and ANTHONY OSBORNE, 49
Paradoxical Connections, ROBERT J. MACG. DAWSON, 31
Patterns in Linear Algebra, GILBERT STRANG, 105
Pi, Euler, and Asymptotic Expansions, J. F. BORWEIN, P. B. BORWEIN, and K. DILCHER, 681
A Pictorial Proof of Uniform Continuity, D. M. BLOOM 250
The Polar Decomposition and a Matrix Inequality, DERMING WANG, 517
Polynomials All of Whose Derivatives Have Integer Roots, C. E. CARROLL, 129
Powers of a Prime Dividing Binomial Coefficients W. J. WONG, 513
The Problem of Calissons, GUY DAVID and CARLOS TOMEI, 429
A Problem of Leo Moser About Repeated Distances on the Sphere, PAUL ERDŐS, DEAN HICKERSON, and J. PACH, 569
Pursuing Analogies Between Differential Equations and Difference Equations, DAVID L. ABRAHAMSON, 827
The Radius of Convergence of Power Series Solutions to Linear Differential Equations, ISOM H. HERRON, 824
Ramanujan, Modular Equations, and Approximations to Pi or How to Compute One Billion Digits of Pi, J. M. BORWEIN, P. B. BORWEIN, and D. H. BAILEY, 201
Reconstructing a Function from Its Set of Tangent Lines, ALAN HORWITZ, 807
Reflection Sequences, N. ALON, I. KRASIKOV, and Y. PERES, 820
A Remark on Divided Differences, E. T. Y. LEE, 618
A Remark on Euclid's Proof of the Infinitude of Primes, JOHN B. COSGRAVE, 339
Rings with Invertible Regular Elements, F. C. LEARY, 924
The Shortest Planar Arc of Width 1, ANI ADHIKARI AND JIM PITMAN, 309
A Simple Estimate of the Error in Linear Approximation, ROBERT M. GETHNER, 522
Simple Inequalities and Old Limits, CHUNG-LIE WANG, 354
Some Recent Geometric Inequalities, FARUK F. ABI-KHUZAM and ARTIN B. BOGHOSSIAN, 576
Sum Zero (mod $n$), Size $n$ Subsets of Integers, CRAIG BAILEY AND R. BRUCE RICHTER, 240
Summability Theory: A Neglected Tool of Analysis, LEE A. RUBEL, 421
A Supplement to I. N. Herstein's Remark on Finite Fields, HARAGAURI N. GUPTA, 733
A Survey of Transcendentally Transcendental Functions, LEE A. RUBEL, 777
Taylor's Theorem Using the Generalized Riemann Integral, H. B. THOMPSON, 346
There Are No Safe Virus Tests, WILLIAM F. DOWLING, 835
Uniqueness of Representation by Trigonometric Series, J. MARSHALL ASH, 873
Unsolved Problems Come of Age, RICHARD K. GUY, 903
The Use of Full Covers in Real Analysis, MICHAEL W. BOTSKO, 328
A Version of Rouché's Theorem for Continuous Functions, A. TSARPALIAS, 911
A Very Short Proof of Stirling's Formula, J. M. PATIN, 41
What Are the Laws of Greed?, JIM PROPP, 334
The White Screen Problem, HERBERT S. WILF, 704

# AUTHOR INDEX


ABI-KHUZAM, FARUK F. and ARTIN B. BOGHOSSIAN, Some Recent Geometric Inequalities, 576

ABRAHAMSON, DAVID L., Pursuing Analogies Between Differential Equations and Difference Equations, 827

ADHIKARI, ANI and JIM PITMAN, The Shortest Planar Arc of Width 1, 309

AGUIRRE, JOSE A. FACENDA, A Note on Taylor's Theorem, 244

ALON, N., I. KRASIKOV, and Y. PERES, Reflection Sequences,

ANDERSON, RICHARD, L. LOVÁSZ, P. SHOR, J. SPENCER, E. TARDOS, and S. WINOGRAD, Disks, Balls, and Walls: Analysis of a Combinatorial Game, 481

ANDREWS, GEORGE E. and R. J. BAXTER, A Motivated Proof of the Rogers-Ramanujan Identities, 401

ARTERBURN, DAVID R. and WILLIAM D. STONE, A Cantor Set of Nonconvergence, 604

ASH, J. MARSHALL, Uniqueness of Representation by Trigonometric Series, 873

BAILEY, CRAIG and R. BRUCE RICHTER, Sum Zero (mod $n$), Size $n$ Subsets of Integers, 240

BAILEY, D. H., J. M. BORWEIN, and P. B. BORWEIN, Ramanujan, Modular Equations, and Approximations to Pi or How to Compute One Billion Digits of Pi, 201

BALDWIN, JOHN T., On Markov Processes in Elementary Mathematics Courses, 147

BAPAT, R. B., Comparing the Spectral Radii of Two Nonnegative Matrices, 137

BASSEIN, RICHARD, An Optimization Problem, 721

BATEMAN, P. T., J. L. SELFRIDGE, and S. S. WAGSTAFF, JR., The New Mersenne Conjecture, 125

BAXTER, R. J. and GEORGE E. ANDREWS, A Motivated Proof of the Rogers-Ramanujan Identities, 401

BEZDEK, K. and R. CONNELLY, Covering Curves by Translates of a Convex Set, 789

BLANK, S. J., NISHAN KRIKORIAN, and DAVID SPRING, A Geometrically Inspired Proof of the Singular Value Composition, 238

BLOOM, D. M., A Pictorial Proof of Uniform Continuity, 250

BOCKMAN, S. F., Generalizing the Formula for Areas of Polygons to Moments, 131

BOGART, KENNETH P., A (Fresh)man Treatment of Determinants, 915

BOGHOSSIAN, ARTIN and FARUK F. ABI KHUZAM, Some Recent Geometric Inequalities, 576

BORWEIN, J. M., P. B. BORWEIN, and K. DILCHER, Pi, Euler, and Asymptotic Expansions, 681

BORWEIN, J. M., P. B. BORWEIN, and D. H. BAILEY, Ramanujan, Modular Equations, and Approximations to Pi or How to Compute One Billion Digits of Pi, 201

BORWEIN, P. B., J. M. BORWEIN, and K. DILCHER, Pi, Euler, and Asymptotic Expansions, 681

BORWEIN, P. B., J. M. BORWEIN, and D. H. BAILEY, Ramanujan, Modular Equations, and Approximations to Pi or How to Compute One Billion Digits of Pi, 201

BOTSKO, MICHAEL W., The Use of Full Covers in Real Analysis, 328

BRADLEY, RICHARD C., An Elementary Treatment of the Radon-Nikodym Derivative, 437

BURTON, MARTHA B., The Effect of Prior Calculus Experience on 'Introductory' College Calculus, 350

BUSENBERG, STAVROS, DAVID FISHER, and MARIO MARTELLI, Minimal Periods of Discrete and Smooth Orbits, 5

CANNON, JAMES W., The Cauchy-Schwarz Inequality: A Geometric Proof, 630

CARROLL, C. E., Polynomials All of Whose Derivatives Have Integer Roots, 129

COHEN, JOEL, Monotone Multiplicative Functions, 512

CONNELLY, R. and K. BEZDEK, Covering Curves by Translates of a Convex Set, 789

COOPER, CURTIS N. and ROBERT E. KENNEDY, Chebyshev's Inequality and Natural Density, 118

CORBALAN, ANTONIO G. and FELIPE CUCKER, An Alternate Proof of the Continuity of the Roots of a Polynomial, 342

COSGRAVE, JOHN B., A Remark on Euclid's Proof of the Infinitude of Primes, 339

CROWE, JOHN and DOMINICK SAMPERI, On Open Maps, 242

CUCKER, FELIPE and ANTONIO G. CORBALAN, An Alternate Proof of the Continuity of the Roots of a Polynomial, 342

CUSICK, LARRY W. and PETER TANNENBAUM, Fixed Points of the Twisted Cyclic Shift Operator, 713

DAVID, GUY and CARLOS TOMEI, The Problem of Calissons, 429

DAWSON, ROBERT J. MACG., Paradoxical Connections, 31

DIACONIS, PERSI and JOSEPH B. KELLER, Fair Dice, 337

DILCHER, K., J. M. BORWEIN, and P. B. BORWEIN, Pi, Euler, and Asymptotic Expansions, 681

DOWLING, WILLIAM F., There Are No Safe Virus Tests, 835

DUEMMEL, JAMES, From Calculus to Number Theory, 140

EDELSBRUNNER, HERBERT and STEVEN SKIENA, On the Number of Furthest Neighbour Pairs in a Point Set, 614

EPPERSON, JAMES F., On the Use of Iteration Methods for Approximating the Natural Logarithm, 831

ERDŐS, PAUL, DEAN HICKERSON, and J. PACH, A Problem of Leo Moser About Repeated Distances on the Sphere, 569

FAY, T. H., The Butterfly Curve, 442
FISHER, DAVID, STAVROS BUSENBERG, and MARIO MARTELLI, Minimal Periods of Discrete and Smooth Orbits, 5
FISHER, DAVID C., The Number of Words of Length $n$ in a Graph Monoid, 610
FULDA, JOSEPH S., Material Implication Revisited, 247
GALLIAN, JOSEPH A, Fifty Years of Putnam Trivia, 711
GAMELIN, T. W. and D. KHAVINSON, The Isoperimetric Inequality and Rational Approximation, 18
GETHNER, ROBERT M., A Simple Estimate of the Error in Linear Approximation, 522
GOODMAN, JACOB E., JÁNOS PACH, and CHEE K. YAP, Mountain Climbing, Ladder Moving, and the Ring-Width of a Polygon, 494
GUPTA, HARAGAURI N., A Supplement to I. N. Herstein's Remark on Finite Fields, 733
GUPTA, NARAIN, On Groups in Which Every Element Has Finite Order, 297
GUY, RICHARD K., Unsolved Problems Come of Age, 903
GUY, RICHARD K., Conway's RATS and Other Reversals, 425
HARTIG, DONALD, On the Differentiation Formula for sin, 252
HERRON, ISOM H., The Radius of Convergence of Power Series Solutions to Linear Differential Equations, 824
HICKERSON, DEAN, PAUL ERDŐS and J. PACH, A Problem of Leo Moser About Repeated Distances on the Sphere, 569
HORWITZ, ALAN, Reconstructing a Function from Its Set of Tangent Lines, 807
HUPERT, LILLIAN E. PETERS and ANNE LEGGETT, On the Square Roots of Infinite Matrices, 34
IRWIN, M. C., Geometry of Continued Fractions, 696
JENSEN, BJARNE S. and MOGENS E. LARSEN, Coupled Linear Differential Equations with Real Coefficients, 729
JOHNSON, STEVE, Factor Rings of Integers, 521
KAPPE, LUISE-CHARLOTTE and BETTE WARREN, An Elementary Test for the Galois Group of a Quartic Polynomial, 133
KASS, SEYMOUR, An Eigenvalue Characterization of the Correlation Coefficient, 910
KELLER, JOSEPH B. and PERSI DIACONIS, Fair Dice, 337
KENNEDY, ROBERT E. and CURTIS N. COOPER, Chebyshev's Inequality and Natural Density, 118
KHAVINSON, D. and T. W. GAMELIN, The Isoperimetric Inequality and Rational Approximation, 18
KRASIKOV, I., N. ALON, and Y. PERES, Reflection Sequences, 820
KRIKORIAN, NISHAN, S. J. BLANK, and DAVID SPRING, A Geometrically Inspired Proof of the Singular Value Composition, 238
LARSEN, MOGENS E. and BJARNE S. JENSEN, Coupled Linear Differential Equations with Real Coefficients, 729
LAWRENCE, J. and G. SIMONS, Equations in Division Rings—A Survey, 220
LEARY, F. C., Rings with Invertible Regular Elements, 924
LEE, E. T. Y., A Remark on Divided Differences, 618
LEGGETT, ANNE and LILLIAN E. PETERS HUPERT, On the Square Roots of Infinite Matrices, 34
LEHNING, HERVÉ, From Experimentation to Proof, 631
LEW, JOHN S. and DONALD A. QUARLES, JR., Optimal Inscribed Polygons in Convex Curves, 886
LEWKOWICZ, R. E., A Characterization of a Class of Composition Operator, 725
LIEBECK, HANS and ANTHONY OSBORNE, Orthogonal Bases of $\mathbb{R}^3$ with Integer Coordinates and Integer Lengths, 49
LIGH, STEVE and JIANG LUH, Direct Sum of $J$-Rings and Zero Rings, 40
LOVÁSZ, LÁSZLÓ, R. ANDERSON, P. SHOR, J. SPENCER, E. TARDOS, and S. WINOGRAD, Disks, Balls, and Walls: Analysis of a Combinatorial Game, 481
LUH, JIANG and STEVE LIGH, Direct Sum of $J$-Rings and Zero Rings, 40
MADDOX, I. J., The Norm of a Linear Function, 434
MADER, ADOLF, Euclidean Model for Euclidean Geometry, 43
MARTELLI, MARIO, STAVROS BUSENBERG, and DAVID FISHER, Minimal Periods of Discrete and Smooth Orbits, 5
McCLEARY, JOHN, How Not to Prove Fermat's Last Theorem, 410
NIEVERGELT, YVES, Estimating the Diffusion of Stock Prices with the HP-28S, 636
OSBORNE, ANTHONY and HANS LIEBECK, Orthogonal Bases of $\mathbb{R}^3$ with Integer Coordinates and Integer Lengths, 49
PACH, J., PAUL ERDŐS, and DEAN HICKERSON, A Problem of Leo Moser About Repeated Distances on the Sphere, 569
PACH, JÁNOS, JACOB E. GOODMAN, and CHEE K. YAP, Mountain Climbing, Ladder Moving, and the Ring-Width of a Polygon, 494
PAKULA, LEWIS, A Note on Venn Diagrams, 38
PAMBUCCIAN, VICTOR, Another Example of an Exotic Function, 913
PATIN, J. M., A Very Short Proof of Stirling's Formula, 41
PERES, Y., N. ALON, and I. KRASIKOV, Reflection Sequences, 820
PITMAN, JIM and ANI ADHIKARI, The Shortest Planar Arc of Width 1, 309
PROPP, JIM, What are the Laws of Greed?, 334
QUARLES, JR., DONALD A. and JOHN S. LEW, Optimal Inscribed Polygons in Convex Curves, 886
RICHTER, R. BRUCE and CRAIG BAILEY, Sum Zero $(\bmod n)$, Size $n$ Subsets of Integers, 240
ROSAY, JEAN-PIERRE and WALTER RUDIN, Arakelian's Approximation Theorem, 432

ROSSET, SHMUEL, Normalized Symmetric Functions, Newton's Inequalities, and a New Set of Stronger Inequalities, 815
RUBEL, LEE A., Summability Theory: A Neglected Tool of Analysis, 421
RUBEL, LEE A., A Survey of Transcendentally Transcendental Functions, 777
RUDIN, WALTER and JEAN-PIERRE ROSAY, Arakelian's Approximation Theorem, 432
RUMMLER, HANSKLAUS, A Natural Interpretation of an Artificial Function, 520
RUSHANAN, JOSEPH J., On the Vandermonde Matrix, 921
SAGHER, YORAM, Counting the Rationals, 823
SAH, CHIH-HAN, A Note on the Row-Reduction Algorithm, 143
SAMPERI, DOMINICK and JOHN CROWE, On Open Maps, 242
SANGWINE-YAGER, J. R., The Missing Boundary of the Blaschke Diagram, 233
SELFRIDGE, J. L., P. T. BATEMAN, and S. S. WAGSTAFF, JR., The New Mersenne Conjecture, 125
SHAW, RONALD and FRED J. YEADON, On $(\mathbf{a} \times \mathbf{b}) \times \mathbf{c}$, 623
SHOR, PETER, R. ANDERSON, L. LOVÁSZ, J. SPENCER, E. TARDOS, and S. WINOGRAD, Disks, Balls, and Walls: Analysis of a Combinatorial Game, 481
SIMONS, G. and J. LAWRENCE, Equations in Division Rings—A Survey, 220
SKIENA, STEVEN and HERBERT EDELSBRUNNER, On the Number of Furthest Neighbour Pairs in a Point Set, 614
SPENCER, JOEL, R. ANDERSON, L. LOVÁSZ, P. SHOR, E. TARDOS, and S. WINOGRAD, Disks, Balls, and Walls: Analysis of a Combinatorial Game, 481
SPRING, DAVID, S. J. BLANK, and NISHAN KRIKORIAN, A Geometrically Inspired Proof of the Singular Value Composition, 238
SPROWS, DAVID J., Irrationals and the Fundamental Theorem of Arithmetic, 732
STOCK, DANIEL L, Merlin's Magic Square Revisited, 608
STONE, WILLIAM D. and DAVID R. ARTERBURN, A Cantor Set of Nonconvergence, 604
STRANG, GILBERT, Patterns in Linear Algebra, 105
TANNENBAUM, PETER and LARRY W. CUSICK, Fixed Points of the Twisted Cyclic Shift Operator, 713
TARDOS, EVA, R. ANDERSON, L. LOVÁSZ, P. SHOR, J. SPENCER, and S. WINOGRAD, Disks, Balls, and Walls: Analysis of a Combinatorial Game, 481
THOMPSON, H. B., Taylor's Theorem Using the Generalized Riemann Integral, 346
THOMPSON, CHRISTOPHER, L., Nonstandard Continuity and Uniform Convergence, 443
THURSTON, HUGH, Can a Graph be Both Continuous and Discontinuous?, 814
TOMEI, CARLOS and GUY DAVID, The Problem of Calissons, 429
TSARPALIAS, A., A Version of Rouché's Theorem for Continuous Functions, 911
VAN DOUWEN, ERIC K., Fubini's Theorem for Null Sets, 718
WAGSTAFF, JR., S. S., J. L. SELFRIDGE, and P. T. BATEMAN, The New Mersenne Conjecture, 125
WANG, DERMING, The Polar Decomposition and a Matrix Inequality, 517
WANG, CHUNG-LIE, Simple Inequalities and Old Limits, 354
WARREN, BETTE and LUISE-CHARLOTTE KAPPE, An Elementary Test for the Galois Group of a Quartic Polynomial, 133
WILANSKY, ALBERT, On the Proof of the Radon-Nikodym Theorem, 441
WILF, HERBERT S., The White Screen Problem, 704
WINOGRAD, SHMUEL, R. ANDERSON, L. LOVÁSZ, P. SHOR, J. SPENCER, and E. TARDOS, Disks, Balls, and Walls: Analysis of a Combinatorial Game, 481
WONG, W. J., Powers of a Prime Dividing Binomial Coefficients, 513
YAP, CHEE K., JACOB E. GOODMAN, and JÁNOS PACH, Mountain Climbing, Ladder Moving, and the Ring-Width of a Polygon, 494
YEADON, FRED J. and RONALD SHAW, On $(\mathbf{a} \times \mathbf{b}) \times \mathbf{c}$, 623
ZEILBERGER, DORON, Kathy O'Hara's Constructive Proof of the Unimodality of the Gaussian Polynomials, 590
ZEILBERGER, DORON, On a Conjecture of R. J. Simpson About Exact Covering Congruences, 243

# SUBJECT INDEX

This index uses the 1985 revision of the AMS's 1980 Mathematics Subject Classification.

**00-XX GENERAL**

A08 Mathematical recreation
YORAM SAGHER 823

**01-XX HISTORY AND BIOGRAPHY**

A99 Miscellaneous topics
JOSEPH A. GALLIAN 711

**03-XX MATHEMATICAL LOGIC AND FOUNDATIONS**

B10 General logic: Pure first-order logic (including many-sorted logic)
J. S. FULDA 247

D10 Recursion theory: Turing machines and related notions.
WILLIAM F. DOWLING 835

H05 Nonstandard models: Infinitesimal analysis in pure mathematics
C. L. THOMPSON 443

**05-XX COMBINATORICS**

A10 Classical combinatorial problems: Factorials, binomial coefficients, combinatorial functions
W. J. WONG 513 DORON ZEILBERGER 590

A15 Classical combinatorial problems: Combinatorial enumeration problems, generating functions
DAVID C. FISHER 610

A19 Classical combinatorial problems: Combinatorial identities
G. E. ANDREWS AND R. J. BAXTER 401

A99 Classic combinatorial problems: None of the above, but in this section
R. ANDERSON, L. LOVÁSZ, P. SHOR, J. SPENCER, E. TARDOS, AND S. WINOGRAD 481 N. ALON, I. KRASIKOV, AND Y. PERES 820

C35 Graph theory: Extremal problems
HERBERT S. WILF 704

**11-XX NUMBER THEORY**

A25 Elementary number theory: Arithmetic functions; related numbers; inversion formulas
J. COHEN 512 P. T. BATEMAN, J. L. SELFRIDGE, AND S. S. WAGSTAFF, JR. 125

A51 Elementary number theory: Factorization; primality
DAVID J. SPROWS 732

A55 Elementary number theory: Continued fractions
M. C. IRWIN 696

B25 Sequences and sets: Arithmetic progressions
J. PROPP 334

C20 Polynomials and matrices: Matrices, determinants
H. LIEBECK AND A. OSBORNE 49

D41 Diophantine equations: Higher degree equations; Fermat's equation
J. MCCLEARY 410

D99 Diophantine equations: None of the above, but in this section
J. DUEMMEL 140

**13-XX COMMUTATIVE RINGS AND ALGEBRAS**

A05 General commutative ring theory: Divisibility
F. C. LEARY 924

M05 Finite commutative rings: Structure
S. JOHNSON 521

**15-XX LINEAR AND MULTILINEAR ALGEBRA; MATRIX THEORY**

-01 Elementary exposition; textbooks
G. STRANG 105

A21 Canonical forms, reductions, classification
C.-H. SAH 143

A36 Matrices of integers
H. LIEBECK AND A. OSBORNE 49

A45 Miscellaneous inequalities involving matrices
D. WANG 517

**16-XX ASSOCIATIVE RINGS AND ALGEBRAS**

A02 Integral domains, unique factorization domains (noncommutative)
DAVID J. SPROWS 732

A04 Noncommutative principal ideal rings, rings with a division algorithm
J. LAWRENCE AND G. SIMONS 220

A39 Skew fields, division rings
J. LAWRENCE AND G. SIMONS 220

A44 Finite rings
S. JOHNSON 521

**20-XX GROUP THEORY AND GENERALIZATIONS**

-02 Advanced exposition (research surveys, monographs, etc.)
N. GUPTA 297

**26-XX REAL FUNCTIONS**

-01 Elementary exposition; textbooks
M. BOTSKO 328 ALAN HORWITZ 807

A06 Functions of one variable: One-variable calculus
H. B. THOMPSON 346 M. B. BURTON 350 C. HARTIG 252

A09 Functions of one variable: Elementary functions
D. HARTIG 252 C.-L. WANG 354

A15 Functions of one variable: continuity and related questions (modulus of continuity, semicontinuity, discontinuities, etc.)
H. RUMMLER 520 D. M. BLOOM 250

A18 Functions of one variable: iteration
DAVID R. ARTERBURN AND WILLIAM D. STONE 604

A30 Functions of one variable: Singular functions, Cantor functions, functions with other special properties
VICTOR PAMBUCCIAN 913

A36 Functions of one variable: Antidifferentiation
H. B. THOMPSON 346

A99 Functions of one variable: None of the above, but in this section
LEE A. RUBEL 777 T. H. FAY 442

B12 Functions of several variables: Calculus of vector functions
J. A. F. AGUIRRE 244

C05 Polynomials, rational functions: polynomials: analytic properties, inequalities, etc.
SHMUEL ROSSET 815

C10 Polynomials, rational functions: Polynomials: location of zeros
A. G. CORBALAN AND F. CUCKER 342

D05 Inequalities: Inequalities for trigonometric functions and polynomials
SHMUEL ROSSET 815

**28-XX MEASURE AND INTEGRATION**

A25 Classical measure theory: Integration with respect to measures and other set functions
R. C. BRADLEY 437 A. WILANSKY 441

A35 Classical measure theory: Measures and integrals in product spaces
ERIC K. VAN DOUWEN 718

**30-XX FUNCTIONS OF A COMPLEX VARIABLE**

-01 Elementary exposition; textbooks
A. TSARPALIAS 911

**34-XX ORDINARY DIFFERENTIAL EQUATIONS**

-01 Elementary exposition; textbooks
HERVÉ LEHNING 631

A30 General theory: Linear equation and systems
BJARNE S. JENSEN AND MOGENS E. LARSEN 729

A50 General theory: Numerical approximation of solutions
HERVÉ LEHNING 631

A99 General Theory: None of the above, but in this section
DAVID L. ABRAHAMSON 827

**39-XX FINITE DIFFERENCES AND FUNCTIONAL EQUATIONS**

A10 Finite Differences: Difference equations
DAVID L. ABRAHAMSON 827

**40-XX SEQUENCES, SERIES, SUMMABILITY**

-01 Elementary exposition; textbooks
L. A. RUBEL 421

A05 Convergence and divergence of infinite limiting processes: Convergence and divergence of series and sequences
DAVID R. ARTERBURN AND WILLIAM D. STONE 604

**41-XX APPROXIMATIONS AND EXPANSIONS**

A30 Approximation by other special function classes
J.-P. ROSAY AND W. RUDIN 432

A80 Remainders in approximation formulas
R. M. GETHNER 522

**42-XX FOURIER ANALYSIS**

-01 Elementary exposition; textbooks
J. MARSHALL ASH 873

**46-XX FUNCTIONAL ANALYSIS**

E15 Linear function spaces and their duals: Banach spaces of continuous, differentiable or analytic spaces
I. J. MADDOX 434

**51-XX GEOMETRY**

-01 Elementary exposition; textbooks
A. ADHIKARI AND J. PITMAN 309 J. E. GOODMAN, JANOS PACH, AND CHEE K. YAP 494
S. BUSENBERG, D. FISHER, AND M. MARTELLI 5
FARUK F. ABI-KHUZAM AND A. B. BOGHOSSIAN 576
T. W. GAMELIN AND D. KHAVINSON 18

K05 Distance geometry: General theory
PAUL ERDÖS, DEAN HICKERSON, AND J. PACH 569

M05 Real and complex geometry: Euclidean geometries
A. MADER 43

M20 Real and complex geometry: Length, area, and volume
P. DIACONIS AND J. B. KELLER 337

**52-XX CONVEX SETS AND RELATED GEOMETRIC TOPICS**

A05 Convex sets without dimension restrictions
K. BEZDEK AND R. CONNELLY 789

**54-XX GENERAL TOPOLOGY**

C05 Maps and general types of spaces defined by maps: Continuous maps
HUGH THURSTON 814

C10 Maps and general types of spaces defined by maps: Special maps: open, closed, perfect, almost open, light, etc.
J. CROWE AND D. SAMPERI 242

**60-XX PROBABILITY THEORY AND STOCHASTIC PROCESSES**

A05 Foundations of probability theory: Axioms; other general questions P. DIACONIS AND J. B. KELLER 337

E15 Distribution theory: Inequalities (Chebyshev, Kolmogrov, etc.)
C. N. COOPER AND R. E. KENNEDY 118
J99 Markov processes: None of the above, but in this section
J. T. BALDWIN 147

**65-XX NUMERICAL ANALYSIS**

-01 Elementary expositions; textbooks
E. T. Y. LEE 618
D10 Numerical approximation: Smoothing, curve fitting
JOHN S. LEW AND DONALD A. QUARLES, JR. 886
D32 Numerical approximation: Quadrature and cubature formulas
E. T. Y. LEE 618
D99 Numerical approximation: None of the above, but in this section
J. M. PATIN 41

**65-XX COMPUTER SCIENCE**

-04 Explicit machine computation and programs (not the theory of computation or programming)
J. M. BORWEIN, P. B. BORWEIN, AND K. DILCHER 681 J. M. BORWEIN, P. B. BORWEIN, AND D. H. BAILEY 201
Q60 Theory of computing: Program verification
WILLIAM F. DOWLING 835

**90-XX ECONOMICS, OPERATIONS RESEARCH, PROGRAMMING GAMES**

A09 Mathematical economics: Portfolio theory and financial economics
YVES NIEVERGELT 636

# REVIEWS BY TITLE

*(Names of authors in ordinary type; those of reviewers in capitals)*

*The Book of Prime Number Records*, Paulo Ribenboim, C. POMERANCE, 663–665
*Algorithms in Combinatorial Geometry*, Herbert Edelsbrunner, J. E. GOODMAN, 457–460
*Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds*, George Pólya and R. C. Read, R. MERRIS, 269–272
*Combinatorics of Finite Sets*, Ian Anderson, D. J. KLEITMAN, 463–464
*A Course in Number Theory and Cryptology*, Neal Koblitz, A. G. KONHEIM, 374–375
*Derive, A Mathematical Assistant, ver. 1.22*, Albert Rich, Joan Rich, and David Stoutemyer, E. A. HERMAN, 948–958
*Dynamical Systems*, V. I. Arnold (ed.), J. HORNSTEIN, 861
*Elementary Number Theory*, Charles Vanden Eynden, E. GROSSWALD, 460–463
*For All Practical Purposes: An Introduction to Contemporary Mathematics*, COMAP, A. GITTLEMAN, 465–467
*Forever Undecided: A Puzzle Guide to Gödel*, Raymond Smullyan, C. SMORYNSKI, 169–172
*From One to Zero: A Universal History of Numbers*, Georges Ifrah, F. SWETZ, 273–274
*Geometric Inequalities*, Yu. D. Burago and V. A. Zalgaller, D. CHAKERIAN, 544–546
*Geometric Measure Theory. A Beginner's Guide*, Frank Morgan, F. J. ALMGREN, JR.
*Geometric Theory of Foliations*, Cesar Camacho and A. L. Neto, A. PHILLIPS, 71–76
*Geometries and Groups*, V. V. Nikulin and I. R. Shafarevich, H. W. GUGGENHEIMER, 370–373
*Invitation to Complex Analysis*, Ralph Boas, G. PIRANIAN, 376–378
*Littlewood's Miscellany*, Bélas Bollobás, R. P. BOAS, 167–169
*Mathematica—A System for Doing Mathematics by Computer*, Wolfram Research, L. S. KROLL, 855
*Mathematical Cryptology for Computer Scientists and Mathematicians*, Wayne Patterson, A. G. KONHEIM, 374–375
*Mathematical Problem Solving*, Alan H. Schoenfeld, S. GALOVICH, 68–71
*Mathematics with Applications*, fourth edition, Margaret L. Lial and C. D. Miller, C. J. OXENRIDER, 537–538
*Native American Mathematics*, Michael P. Closs (ed.), J. V. RAUFF, 662–663
*The Pólya Picture Album—Encounters of a Mathematician*, Gerald A. Alexanderson (ed.), F. HARARY, 750–
*Principles of Computer Science*, M. Sandra Carberry, A. T. Cohen, and H. M. Khalil, S. E. SELTZER, 378–380
*The Shape of Space*, Jeffrey R. Weeks, A. H. DURFEE, 660–662
*Sphere Packings, Lattices, and Groups*, J. H. Conway and N. J. A. Sloane, H. S. M. COXETER, 538–544
*Writing Mathematics Well*, Leonard Gillman, P. D. LAX, 380–381

# REVIEWS BY AUTHOR

*(Names of authors in ordinary type; those of reviewers in capitals)*


Alexanderson (ed.), Gerald A., *The Pólya Picture Album —Encounters of a Mathematician*. F. HARARY, 750–

Anderson, Ian, *Combinatorics of Finite Sets*, D. J. KLEITMAN, 463–464

Arnold (ed.), V. I., *Dynamical Systems*, J. HORNSTEIN, 861

Boas, Ralph, *Invitation to Complex Analysis*, G. PIRANIAN, 376–378

Bollobás, Bélas, *Littlewood's Miscellany*, R. P. BOAS, 167–169

Burago, Yu. D. and V. A. Zalgaller, *Geometric Inequalities*, D. CHAKERIAN, 544–546

Camacho, Cesar and A. L. Neto, *Geometric Theory of Foliations*, A. PHILLIPS, 71–76

Carberry, M. Sandra, A. T. Cohen, and H. M. Khalil, *Princples of Computer Science*, S. E. SELTZER, 378–380

Closs (ed.), Michael P., *Native American Mathematics*, J. V. RAUFF, 662–663

COMAP, *For All Practical Purposes: An Introduction to Contemporary Mathematics*, A. GITTLEMAN, 465–467

Cohen, A. Toni, M. S. Carberry, and H. M. Khalil, *Principles of Computer Science*, S. E. SELTZER, 378–380

Conway, J. H. and N. J. A. Sloane, *Sphere Packings, Lattices, and Groups*, H. S. M. COXETER, 538–544

Edelsbrunner, Herbert, *Algorithms in Combinatorial Geometry*, J. E. GOODMAN, 457–460

Gillman, Leonard, *Writing Mathematics Well*, P. D. LAX, 380–381

Ifrah, Georges, *From One to Zero: A Universal History of Numbers*, F. SWETZ, 273–274

Khalil, Hatem M., M. S. Carberry, and A. T. Cohen, *Principles of Computer Science*, S. SELTZER, 378–380

Koblitz, Neal, *A Course in Number Theory and Cryptology*, A. G. KONHEIM, 374–375

Lial, Margaret L. and C. D. Miller, *Mathematics with Applications*, fourth edition, C. J. OXENRIDER, 537–538

Miller, Charles D. and M. L. Lial, *Mathematics with Applications*, fourth edition, C. J. OXENRIDER, 537–538

Morgan, Frank, *Geometric Measure Theory. A Beginner's Guide*, F. J. ALMGREN, JR.

Neto, Alcides Lins and C. Camacho, *Geometric Theory of Foliations*, A. PHILLIPS, 71–76

Nikulin, V. V. and I. R. Shafarevich, *Geometries and Groups*, H. W. GUGGENHEIMER, 370–373

Patterson, Wayne, *Mathematical Cryptology for Computer Scientists and Mathematicians*, A. G. KONHEIM, 374–375

Pólya, George and R. C. Read, *Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds*, R. MERRIS, 269–272

Read, R. C. and G. Pólya, *Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds*, R. MERRIS, 269–272

Ribenboim, Paulo, *The Book of Prime Number Records*, C. POMERANCE, 663–665

Rich, Albert, Joan Rich, and David Stoutemyer, *Derive, A Mathematical Assistant, ver. 1.22*, E. A. HERMAN, 948–958

Rich, Joan, Albert Rich, and David Stoutemyer, *Derive, A Mathematical Assistant, ver. 1.22*, E. A. HERMAN, 948–958

Schoenfeld, Alan, H., *Mathematical Problem Solving*, S. GALOVICH, 68–71

Shafarevich, I. R. and V. V. Nikulin, *Geometries and Groups*, H. W. GUGGENHEIMER, 370–373

Sloane, N. J. A. and J. H. Conway, *Sphere Packings, Lattices, and Groups*, H. S. M. COXETER, 538–544

Smullyan, Raymond, *Forever Undecided: A Puzzle Guide to Gödel*, C. SMORYNSKI, 169–172

Stoutemyer, David, Albert Rich, and Joan Rich, *Derive, A Mathematical Assistant, ver. 1.22*, E. A. HERMAN, 948–958

Vanden Eynden, Charles, *Elementary Number Theory*, E. GROSSWALD, 460–463

Weeks, Jeffrey R., *The Shape of Space*, A. H. DURFEE, 660–662

Wolfram Research, *Mathematica—A System for Doing Mathematics by Computer*, L. S. KROLL,

Zalgaller, V. A. and Yu. D. Burago, *Geometric Inequalities*, D. CHAKERIAN, 544–546


# 1989 PROBLEMS PROPOSED


Adler, Irving 642

Alex, Leo J. 927

Alexandrov, V. A. 838, 927

Anderson, Oliver D. 652

Andrews, Peter 154

Balazard, Michel 165, 446

Baloglou, George 445
Bege, Antal 253
Bencharit, Um 838
Blakley, Robert 155
Bowman, Doug 743
Brown, Kevin 65
Bulman-Fleming, Sydney 155
Chao, Wu Wei 927
Chatterji, S. D. 265
Clark, Dean S. 735
Comtet, Louis 524
Cucurezeanu, Ion 154
Dairbekov, N. S. 838
Daneshgar, Amir 253
Das, Sajal K. 454
de Doelder, P. J. 366
Delany, Jim 54, 254
Diamond, H. G. 838
Dlab, Vlastimil 735
Duran, Antonio 641
Edgar, Hugh M. W. 942
Ehrhart, E. 533
El-Hayek, C.
Erdös, Paul 356, 837
Evans, Ronald J. 66
Ferraro, Peter J. 253
Ferrer, Jesus 65
Gersten, Stephen M. 165
Gessel, Ira 652
Gleason, Andrew M. 734
Goodman, A. W. 744
Griffith, P. A. 264
Glitzmann, P. 642
Hajek, O. 744
Hensley, Douglas 155
Hildebrand, Adolf 165
Khare, C. B. 253
Klamkin, M. S. 55, 154, 356, 641
Knuth, Donald E. 54, 154, 525
Lau, Kee-Wai
Lupas, Alexandru 357
MacKinnon, Nick 366, 525
Mamer, John 524
MacCarthy, Justin G. 846
McDowell, Kenneth 155
Mercier, Armel 942
Meyerson, Mark D. 357
Miller, Allen R. 454
Moen, Courtney 642
Mohar, B. 642
Montgomery, Hugh L. 54
Murase, Ichiro 265
Pach, J. 642
Plambeck, Thane 357
Pollack, R. 642
Pomerance, Carl 455
Propp, James G. 446
Rabinowitz, Stanley 445, 524
Ratering, Steven 928
Reid, Michael 533
Richter, R. Bruce 357
Rogers, Douglas 734
Rotman, J. J. 264
Rudin, Walter 254, 366, 445, 524, 641
Salehi, Ebrahim 846
Schilling, Kenneth 524
Senum, G. I. 838
Shapiro, Daniel B. 652
Shapiro, Lou 734
Singer, D. 744
Smith, Bruce K. 446
Spira, Robert 356
Stanley, Richard 734
Steele, J. Michael 734
Suranyi, Janos 356
Szekely, Laszlo A. 837
Ullman, Daniel 928
Vaaler, Jeffrey 54
Vanden Eynden, Charles 446
Vialetto, Dante 55
Vince, A. 942
Wang, Edward T. H. 154
Wardlaw, William P. 928
White, Tad 743
Wilansky, Albert 155
Wilker, J. B. 55
Wimp, Jet 838
Witkowski, Alfred 735

## 1989 PROBLEMS SOLVED

Agnew, Robert A. 364
Andrews, George 852
Askey, Richard 653
Bager, Anders 928
Bertram, B. S. 531
Bishop, Richard L. 528
Bloom, David M. 737
Bondesen, Aage 254
Broline, Duane
Callan, David 57, 66, 532, 651, 933
Cavaretta, A. S. 744
Clark, Dean S. 645
Cooper, Curtis 363
David, Karl 530
Delany, Jim 742

Diamond, Harold G. 369
Dodd, Fred 155, 929
Dowling, Roy 844
Driscoll, Richard J. 657
Eddy, R. H. 527
Egerland, W. O. 838
Falkowitz, M. 527
Fee, Gregg 642
Fieldsteel, Adam 650
Filaseta, Michael 166
Foster, Lorraine L. 451
Gagola, Stephen M. 453, 845
Georghiou, C. 58
Gessel, Ira 364
Goodman, T. N. T. 744
Griffin, Peter 847
Hertz, Ellen 651
Herzog, J. 942
Holshouser, A. L. 64
Ilacqua, Paul 357
Iny, David 525
Jager, Thomas 535
Jagers, A. A. 260
Jeurissen, R. H. 59, 367
Johnson, Bruce R. 162
Johnson, Diane 844
Jones, Lenny 361
King, L. R. 64
Klein, B. G. 64
Levine, Leo M. 740
Lewis, Timothy S. 164
Loo, Joseph 451
Lord, N. J. 155
Lossers, O. P. 157, 259, 260, 261, 265, 359, 361, 365, 647, 745, 841
Lucier, Bradley 448
Mallows, C. L. 529
Marco, Ricardo Perez 649
Marston, Helen M. 647
Mattics, L. E. 156, 455, 739, 929
Mauldon, J. G. 256
Maxsein, T. 942
McGee, Ian 655
Monsky, Paul 258
Neuenschwander, Daniel 55
Passell, Nicholas 842
Peck, G. W. 360
Pedersen, Allan 649
Philipp, Stan 365, 747
Pinkham, R. S. 359
Poonen, Bjorn 55
Powers, R. G. 64
Riese, Adams 161, 530, 934
Roberts, Brooks 740
Rousseau, Cecil 655
Rubel, L. A. 266
Rudin, Walter 934
Ruehr, O. G. 531
Ruehr, Otto 653
Schilling, Kenneth 263
Schoen, Carl 62, 163
Schutt, R. 533
Secrest, David B. 158
Seiffert, Heinz-Jurgen 850
Selvaraj, C. R. 744
Shafer, Robert 55
Shapiro, Lou 535, 651
Smith, P. R. 942
Spaulding, Raymond E. 736
Staley, Patrick A. 362
Steelman, John Henry 449, 452, 527, 843
Students of the 1987 Mathematical Olympiad Program 262
Tanaka, Yokichi 451
Thurston, Louis 651
Tyler, Douglas B. 263
University of South Alabama Problem Group 63, 261
Valk, G. W. 846
Wallace, K. D. 64
Ward, John T. 941
Wells, David M. 162, 743
West, Douglas B. 60
Yeung, Hang-Fai 933
Zakeri, Saeid 741

## 1989 SOLUTIONS

E 2938 447, 937
E 2982 525
E 2983 642
E 2986 357
E 3008 155
E 3023 254
E 3099 937
E 3122 55
E 3138 928
E 3153 933
E 3157 939
E 3162 258
E 3163 57
E 3168 58
E 3170 359
E 3171 645
E 3172 939
E 3175 59
E 3177 647
E 3181 156
E 3182 934
E 3183 157
E 3185 735
E 3186 736
E 3187 60
E 3189 259
E 3191 161
E 3192 526
E 3193 738
E 3194 260
E 3195 527
E 3199 648

E 3206 161, 939
E 3208 62
E 3209 261
E 3212 63
E 3213 162
E 3214 739
E 3215 63
E 3216 361
E 3217 64
E 3218 838
E 3219 163, 940
E 3221 841
E 3223 164
E 3226 262
E 3227 447
E 3228 448
E 3230 263
E 3231 449
E 3232 528
E 3233 361
E 3234 362
E 3235 363
E 3237 364
E 3238 451
E 3240 529
E 3241 452
E 3242 365
E 3243 530
E 3244 531
E 3245 532
E 3246 935
E 3247 452
E 3248 649
E 3249 741
E 3250 453
E 3252 650
E 3258 651
E 3261 742
E 3266 743
E 3268 842
E 3269 843
E 3270 844
E 3271 845
E 3272 845

6477 945
6528 947
6531 367
6540 165
6541 653
6542 66
6543 744
6544 265
6545 266
6547 368
6548 533
6549 455
6550 655
6551 657
6553 535
6554 745
6556 847
6560 747
6561 849
6562 852
6563 942

# 1988 PROBLEMS PROPOSED

Albu, Toma 131
Allen, Charles S. 963
Alsina, C. 872
Andrews, P. 350
Appel, K. I. 555
Bagchi, B. 665
Batnik, P. A. 774
Beebee, J. 654
Blom, Gunnar 953
Bloom, David M. 654, 763
Bondesson, Lennart 131
Book, D. L. 461
Borosh, I. 264
Boyd, C. 351
Bulman-Fleming, Sydney 258
Cade, John J. 763
Carmargo, S. 264
Castrellon, W. 264
Cohoon, David K. 773
Dou, Jordi 350
Edgar, Gerald A. 150
Erdös, Paul 51, 259, 762
Ferraro, Peter J. 655
Fine, N. J. 774
Friedlen, D. M. 456
Frink, Orrin 555, 762
Gerber, H. U. 665
Goffinet, Daniel 873
Golomb, Solomon W. 51, 873
Gonciulea, N. 259
Goodman, A. W. 953
Griggs, J. R. 872
Gupta, H. Das 954
Hayes, B. 456
Hensley, Douglas 264
Ierley, G. 359
Isbell, John 259
Janous, Walther 351, 963
Jockusch, C. G. 555
Keane, Joseph 873
Keane, J. C. 554
Kimberling, Clark 131
Klavzar, S. 655
Knuth, Donald E. 456, 774
Laffey, T. J. 555
Laub, Moshe 559, 880
Laugwitz, Detlef 351
Letac, Gerard 461
Lindstrom, Bernt 554
Lupas, Alexandru 880
MacHale, D. 555
MacKinnon, Nick 51
Marcus, Marvin 954
Masharam, D. 457
Milosevic, D. M. 762
Misra, G. 665
Newman, D. 555
Nicholson, W. K. 763
Noll, W. 554
Parker, E. T. 880
Patruno, Gregg 873
Pelling, M. J. 554
Pesce, Claire 954
Petkovsek, M. 655
Pinch, R. G. E. 351
Rabinowitz, Stanley 352, 655
Reznick, Bruce A. 51
Rodriquez, R. S. 50

Rubel, Lee A. 51, 60, 150, 359
Sakmar, I. A. 259
Sastry, N. S. N. 665
Sauer, T. 654
Sherwood, H. 50
Singmaster, D. 258
Spingarn, Jonathan E. 150
Spiro, C. A. 60
Stanley, R. 559
Stefanov, S. T. 666
Stone, A. H. 457
Subi, C. 456
Trench, W. F. 50
Turner, J. C. 456
Vardi, Ilan 774, 963
Wang, Edward T. H. 258, 350
Waterhouse, William C. 763
Weinstein, G. 60
West, Douglas B. 872
Wiedeman, D. H. 872
Winslow, David G. 132
Yamout, Jihad 954
Young, Robert M. 953

# 1988 PROBLEMS SOLVED

Andersen, K. F. 138
Arnold, G. 661
Askey, R. 873
Behrendt, G. 558
Bent, S. W. 556
Binz, J. C. 133
Bondesen, Aage 565
Borwein, Jonathan M. 557
Bowron, Mark 141
Boyadzhiev, K. 58
Bu, Q.-Y. 666
Callan, David 558, 770, 875
Cater, Frank S. 776
Cohen, J. M. 874
Cseh, L. 656
Cull, P. 558
Deutsch, E. 873
Dou, Jordi 135
Egerland, W. O. 140, 354
Erdös, P. 878
Estes, D. 557
Ferrer, J. 353
Fink, A. M. 134
Fukuta, Jiro 138, 139, 660
Georghiou, C. 963
Gessel, Ira 564
Grunbaum, K. 357
Hajek, O. 667
Hansen, C. E. 354
Hengsgen 664
Hook, Jay 156
Isaacs, I. Martin 562
Israel, Robert B. 266, 360, 560
Jagers, A. A. 53, 262, 769
Jodeit, Max 134
Kappus, Hans 58
Kay, David C. 155
Klamkin, Murray S. 358, 659
Knuth, Donald E. 662
Konecny, Vaclav 661
Lambert, W. M. 260
Lanski, C. 557
Laugwitz, D. 459
Lau, Kee-Wai 355
Leech, J. 764
Lindstrom, Bernt 959
Lossers, O. P. 53, 137, 261, 658, 663, 668, 669, 772, 875, 876, 877, 957, 961
Lupas, Alexandru 264
Mallows, C. L. 157
Marco, R. P. 262
Mattics, L. E. 771, 773, 960
Mauldon, J. G. 55
McCoy, T. L. 462
Meir, Amram 959
Mesihovic, Behdzet 149
Nakassis, Magda P. 767
Newcomb, William A. 460
Pach, Janos 765
Pambuccian, Victor 658, 774
Park, K. K. 354
Pedersen, Allan 954
Pedersen, S. 357
Peele, R. 876
Pelling, M. J. 139, 667, 777
Poonen, Bjorn 56, 457
Propp, James 133
Richman, David R. 661
Rogers, C. A. 765
San Bernardino Problem Solving Group 148
Schilling, Kenneth 151, 359
Schuchat, Alan 140
Schwenk, Allen J. 352
Secrest, David B. 461, 873
Shafer, Robert E. 769
Sivakumar, N. 52
Steelman, John Henry 881
Strauss, N. 54

Students in the 1987 Mathematical Olympiad Program 879
Tomsky, J. 665
Turnwald, G. 356
Tzermias, P. 764
University of South Alabama Problem Group 563
Zwier, Paul J. 662

## 1987 PROBLEMS PROPOSED

Andrews, George E. 1011
Arias-de-Reyna, Juan 469
Batnik, P. A. 681
Belinfante, Johan G. F. 995
Bennett, Grahame 371
Berg, I. D. 456
Bishop, R. L. 456
Bloom, D. M. 548
Bode, Dieter 995
Butzer, P. L. 469
Cater, F. S. 81, 195
Cavaretta, A. S. 387
Chang, Geng-Zhe 456
Chico Problem Group 373
Choe, Boo Rim 81, 799
Christophe, L. Matthew 559
Cipu, Mihai 877
Clark, Dean S. 372
Cohoon, David K. 787
Cseh, Lâszló 548
Deaconescu, Marian 877
Diamond, Harold G. 456
Dou, Jordi 876
Egerland, W. O. 549
Elkies, Noam D. 877
Erdös, Paul 301, 372
Fine, N. J. 800
Fisher, D. 181
Froemke, Jon 549
Gessel, Ira 181, 372
Goffinet, Daniel 387
Grossman, Jerrold W. 549
Guelicher, Herbert 301, 876
Helstrom, Carl W. 304
Isaacs, I. Martin 300
Janous, Walther 559
Johnson, Charles R. 877
Kimberling, Clark 884
Klamkin, Murray S. 71, 996
Klein, Benjamin G. 885
Krishna, Ambati Jaya 1010
Kronheimer, Erwin 680
Kuipers, L. 300
Laub, Moshe 694, 800
Layman, John 885
Lenard, Andrew 181
Lipscomb, Stephen L. 72
Lucier, Bradley 787
Lupas, Alexandru 681
Luttmann, Rick 877
Lutzky, M. 181
Macdonald, Ian D. 695
Mallows, C. L. 996
Martelli, Mario 181
McMillan, Brockway 81
Merényi, Imre 548
Miles, J. B. 996
Miller, William 787
Montgomery, Hugh L. 372
Odlyzko, Andrew M. 386
Pfiefer, Richard E. 549
Poonen, Bjorn 786
Popescu, Calin P. 71
Powell, Barry 884
Rao, A. Murali Mohan 1010
Rao, Gomathi S. 1010
Răuţu, Gheorghe 181
Rawsthorne, Daniel 680
Reznick, Bruce A. 457, 681
Rotman, Joseph 195
Rubel, Lee A. 300, 469, 559, 996
Sakmar, I. A. 71
Schoenberg, I. J. 301
Schwenk, Allen J. 72
Selvaraj, C. R. 387
Seiffert, Hans-Jurgen 1011
Smith, Paul A. 457
Solomon, Vasanth B. 181
Song-Qing, Mo 303
Spaulding, Raymond E. 71
Spiro, C. A. 303
Stark, E. L. 469
Stevens, Gary E. 787
Tissier, A. 694
Tomescu, Ioan 72
Tóth, Lâszló 457
Wagon, Stan 786
Wene, Gregory P. 996
Witte, Dave 300
Wolkowicz, Gail 877
Wolkowicz, Henry 877
Wu, Pei Yuan 194
Zhang, Guo Qiang 548
Zhu, J. 457

# 1987 PROBLEMS SOLVED

Barger, S. F. 552
Bennett, Colin 192
Binz, J. C. 376, 378
Boghossian, Artin 997
Bolis, T. S. 458
Bondesen, Aage 373
Borwein, Peter 390
Brenner, J. L. 183
Breusch, Robert 73
Broline, Duane 803
Chernoff, Paul R. 467
The Chico Problem Group 80
Cooke, Roger 473
Dade, Everett C. 696
Delmer, Thomas N. 800
Dixon, David J. 880
Dixon, Edmond D. 882
Edgar, G. A. 695
Ehrlich, Gertrude 190
Eisenberg, Bennett 78
Enkers, D. 190
Eves, Howard 80
Fong, Pamela Y. C. 795
Franco, Z. 556
Freeman, J. M. 188
Fukuta, Jiro 188, 465
Gagola, Steven M. 381
Gardner, C. S. 555
Garfunkel, Jack 80
Garstang, R. H. 557
Gasper, George 200
Georghiou, C. 1006
Gerbic, James 558
Gilbert, William J. 550
Goldstein, Richard 798
Harper, Paul 685
Hensley, Doug 304
Hertz, Ellen 470
Hickman, James 686
Hopkins, Glenn 686
Jagers, A. A. 85, 185, 195
Janous, Walther 686
Jepsen, Charles H. 1003
Jinghuang, Tian 302
Kanetkar, S. V. 462
Kim, Hong Oh 459, 682
Kistner, James E. 880
Klamkin, Murray S. 384
Klein, Edwin M. 385
Knuth, Donald E. 189, 376
Kuipers, L. 690
Lanski, Charles 469
Lau, Kee-Wai 883
Leuck, David H. 187
Locke, S. C. 188
Lossers, O. P. 74, 75, 78, 82, 383, 461, 464, 683, 684, 691, 885, 1000, 1001, 1005, 1007
Manes, David E. 306
Marcus, Dan 379
Mattics, L. E. 182
Meir, A. 384
Mellissen, J. B. M. 1002
Meyerson, Mark D. 793
Nair, Leila 80
Nelsen, Roger B. 467
Neuenschwander, Daniel 386
Neuts, Marcel F. 472
Newcomb, William A. 83, 196
Nichols, Warren 466
Niederhausen, H. 188
Pambuccian, Victor 687, 794
Patruno, Gregg 1012
Pelling, M. J. 1018
Poonen, Bjorn 693, 880
Rangarajan, S. K. 378
Richman, D. 549
Richter, Bruce 793
Rhoades, Billy E. 688
Rudin, Walter 553
Salamin, Eugene 388
San Bernardino Problem Solving Group 802
Schwenk, Allen J. 885
Silberger, D. M. 183
Smith, Paul R. 697
Smyth, C. J. 552
Tyler, Douglas B. 550
University of South Alabama Problem Group 186, 383
Villani, A. 682
Vitale, R. A. 1015
Western Maryland College Problems Group 555, 800
Weston, Steven R. 75
West, Douglas B. 997
Willekens, E. 1018
Williams, Gordon 377
Wu, Pei Yuan 878
Yeung, Hang-Fai 381
Young, Robert L. 77
Zwicker, William S. 1016

THE AMERICAN
# MATHEMATICAL MONTHLY

Volume 96, Number 10 December 1989

## Contents

(ISSN 0002-9890)

ARTICLES

Uniqueness of Representation by Trigonometric Series . . . . . J. MARSHALL ASH 873

Optimal Inscribed Polygons in Convex Curves . . . . . JOHN S. LEW AND DONALD A. QUARLES, JR. 886

UNSOLVED PROBLEMS

Unsolved Problems Come of Age . . . . . RICHARD K. GUY 903

NOTES

An Eigenvalue Characterization of the Correlation Coefficient . . . . . SEYMOUR KASS 910

A Version of Rouché's Theorem for Continuous Functions . . . . . A. TSARPALIAS 911

Another Example of an Exotic Function . . . . . VICTOR PAMBUCCIAN 913

THE TEACHING OF MATHEMATICS

A Fresh(man) Treatment of Determinants . . . . . KENNETH P. BOGART 915

On the Vandermonde Matrix . . . . . JOSEPH J. RUSHANAN 921

Rings With Invertible Regular Elements . . . . . F. C. LEARY 924

PROBLEMS AND SOLUTIONS

Elementary Problems and Solutions . . . . . 927

Advanced Problems and Solutions . . . . . 942

REVIEWS

Derive, A Mathematical Assistant, ver. 1.22 by Albert Rich, Joan Rich, and David Stoutemyer . . . . . EUGENE A. HERMAN 948

TELEGRAPHIC REVIEWS . . . . . 959

INDEX TO VOLUME 96 . . . . . 965

## NOTICE TO AUTHORS

See statement of editorial policy (volume 89, p. 3). Follow the format in current issues. Put your full mailing address in a line at the head of the paper, following the title and the author's name. Send three copies of the manuscript, legibly typewritten on only one side of the paper, double-spaced with wide margins, and keep one as protection against loss. Prepare illustrations carefully, on separate sheets of paper in black ink, the original without lettering and two copies with lettering added. Rough copies of the illustrations should be included in the manuscript at the appropriate places. Supply the full five-symbol Mathematics Subject Classification number, as described in Mathematical Reviews, 1985 and later. (This is necessary for indexing purposes.)

All manuscripts whose length is more than 7 manuscript pages should be sent to HERBERT S. WILF, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104. Shorter articles about the teaching of mathematics should be sent to either MELVIN HENRIKSEN (linear algebra, algebra, differential equations), Department of Mathematics, Harvey Mudd College, Claremont, CA 91711, or STAN WAGON (calculus, analysis, number theory, discrete mathematics, statistics), Department of Mathematics, Smith College, Northampton, MA 01063. Other shorter articles should be submitted as follows: in algebra, discrete mathematics, or probability, to RODICA SIMION, Department of Mathematics, George Washington University, Washington, DC 20052; in geometry or topology to DENNIS DETURCK, Department of Mathematics, University of Pennsylvania, Philadelphia, PA 19104; in analysis to DAVID J. HALLENBECK, Department of Mathematical Sciences, University of Delaware, Newark, DE 19716. Papers in other fields may be sent to any of the editors; however, do not send a paper to more than one editor. Proposed problems or solutions should be sent to PAUL T. BATEMAN, MONTHLY Problems, Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801; unsolved problems to RICHARD GUY, Department of Mathematics and Statistics, University of Calgary, Alberta, Canada T2N 1N4.




Reprint Permission: MARCIA P. SWARD, Executive Director, Mathematical Association of America, 1529 Eighteenth Street, N W., Washington, DC 20036.

Advertising Correspondence. Ms. ELAINE PEDREIRA, Advertising Manager, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036.

Subscription correspondence, changes of address, and inquiries about nondelivered or defective issues: Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20036

Microfilm Editions: University Microfilms International, Serial Bid Coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W , Washington, DC 20036 and Montpelier, VT.

The annual subscription price of $26 for the AMERICAN MATHEMATICAL MONTHLY to an individual member of the Association is included as part of the annual dues. (Annual dues for regular members, exclusive of annual subscription prices for MAA journals, are $29 Student, unemployed and emeritus members receive a 50% discount; new members receive a 30% dues discount for the first two years of membership.) The nonmember/library subscription price is $90 per year.





Second class postage paid at Washington, DC, and additional mailing offices.

Postmaster: Send address changes to the AMERICAN MATHEMATICAL MONTHLY, Membership/Subscriptions Department, Mathematical Association of America, 1529 Eighteenth Street, N.W., Washington, DC 20077-9564.

PRINTED IN THE UNITED STATES OF AMERICA

# Uniqueness of Representation by Trigonometric Series

J. Marshall Ash[1], *DePaul University, Chicago, IL 60614*

J. Marshall Ash received his S.B., S.M., and Ph.D. degrees at the University of Chicago where he was a student of Antoni Zygmund. He was a Joseph Fels Ritt instructor at Columbia University from 1966 to 1969. Since then he has been on the DePaul University faculty, where he became a full professor in 1974 and chaired from 1985 to 1987. In 1977 he was visiting professor at Stanford University. His research interests have centered on multiple trigonometric series, singular integrals, and real analysis, with forays into functional and numerical analysis.

*Abstract.* In 1870 Georg Cantor proved that a $2\pi$ periodic complex valued function of a real variable coincides with the values of at most one trigonometric series. We present his proof and then survey some of the many one dimensional generalizations and extensions of Cantor's theorem. We also survey the situation in higher dimensions, where a great deal less is known.

**1. Cantor's uniqueness theorem.** In 1870 Cantor proved

Theorem C (Cantor [**5**]). *If, for every real number* $x$

$$\lim_{N\to\infty}\sum_{n=-N}^{N} c_n e^{inx} = 0,$$

*then all the complex numbers* $c_n$, $n = 0, 1, -1, 2, -2, \dots$ *are zero.*

This is called a uniqueness theorem because it has as an immediate corollary the fact that a $2\pi$ periodic complex valued function of a real variable coincides with the values of at most one trigonometric series. (Proof: Suppose $\Sigma a_n e^{inx} = \Sigma b_n e^{inx}$ for all $x$. Form the difference series $\Sigma(a_n - b_n)e^{inx}$ and apply Cantor's theorem.)

This theorem is remarkable on two counts. Cantor's formulation of the problem in such a clear, decisive manner was a major mathematical event, given the point of view prevailing among his contemporaries.[2] Equally enjoyable to behold is the rapid resolution that we will now sketch.

Cantor's theorem is relatively easy to prove, if, as Cantor did, you have studied Riemann's brilliant idea of associating to a general trigonometric series $T := \Sigma c_n e^{inx}$, the formal second integral, namely, $F(x) := \Sigma_{n\neq 0}(c_n/(in)^2)e^{inx} + c_0(x^2/2)$. For some interesting remarks on the importance of this idea, see the very enjoyable survey article of Zygmund [**27**]. Define the second Schwarz derivative $D$ of a

[1]The research presented here was supported in part by a grant from the University Research Council of DePaul University.

The author is grateful to one referee for proposing an expanded treatment of multiple trigonometric series and to the other referee for making careful corrections and adding some historical remarks.

[2]In the eighteenth century, physicists just "did" Fourier series (often quite successfully) without worrying about convergence very much at all. When doubts about convergence began to arise in the nineteenth century, the first attempts at rigor were rather heavy handed. See Dauben [**9**, pp. 6–31] for an interesting discussion of the historical context.

function $G(x)$ by

$$DG(x) := \lim_{h\to 0} \frac{G(x+h) - 2G(x) + G(x-h)}{h^2}.$$

The steps of the proof are:

1. Since $T$ converges everywhere, it is immediate that for every value of $x$, $c_n e^{inx} + c_{-n}e^{-inx} \to 0$ as $n \to \infty$. By the Cantor-Lebesgue theorem, $|c_n| + |c_{-n}| \to 0$ as $n \to \infty$. Appendix 1 gives Cantor's weak but easy version of this. (For a statement of the more powerful Cantor-Lebesgue theorem see the survey article by Roger Cooke [**8**]. The proof given there is much shorter than the one in Appendix 1, but requires some of the machinery of modern analysis.)

2. By the Weierstrass $M$-test from

$$\left| \frac{c_n e^{inx}}{(in)^2} + \frac{c_{-n}e^{-inx}}{(-in)^2} \right| \leqslant \frac{\sup(|c_n| + |c_{-n}|)}{n^2}$$

it follows that $F(x) - c_0(x^2/2)$ is a continuous function and that $F$ is the uniform limit of its partial sums. (See Theorems 25.7 and 24.3 in [**14**] for the $M$-test.)

3. An important result of Schwarz's states that if $G$ is continuous and $DG(x) = 0$ for all $x$, then $G$ is a linear function. (See [**5**, pp. 82–83].) Before presenting a proof of this, Cantor remarks that Schwarz mailed this result to him from Zürich.) We give a proof in Appendix 2.

4. Since

$$\frac{e^{i(x+h)} - 2e^{ix} + e^{i(x-h)}}{h^2} = -e^{ix}\left( \frac{\sin\frac{h}{2}}{\frac{h}{2}} \right)^2$$

(Check this.), we have

$$\frac{F(x+h) - 2F(x) + F(x-h)}{h^2} = c_0 + \sum_{n\neq 0} c_n e^{inx} \left( \frac{\sin\frac{nh}{2}}{\frac{nh}{2}} \right)^2.$$

From $a_0 + \sum_{n=1}^{\infty} a_n = 0$ it follows that $\lim_{k\to 0} a_0 + \sum_{n=1}^{\infty} a_n(\sin nk/nk)^2 = 0$. See Appendix 3 for Riemann's summation by parts proof of this.

Hence $DF(x) = 0$ so $F(x) = \alpha x + \beta$ for some $\alpha$ and $\beta$.

5. The right side of the equation

$$c_0\frac{x^2}{2} + \alpha x + \beta = \sum \frac{c_n}{(in)^2} e^{inx}$$

is bounded. (From 2. above it is continuous, hence bounded on $[0, 2\pi]$, and hence bounded everywhere by periodicity.) Letting $x \to \infty$ twice first shows $c_0 = 0$ and then $\alpha = 0$.

6. From the observation made in step 2 above, we see that the sequence

$$s_N(x) := -\beta + \sum_{0<|n|\leqslant N} \frac{c_n}{(in)^2} e^{inx}$$

converges uniformly to 0. But for each $n \neq 0$,

$$c_n = \frac{(in)^2}{2\pi} \int_0^{2\pi} s_N(x) e^{-inx}\, dx$$

for all $N \geqslant n$ by the orthonormality of $\{e^{inx}/\sqrt{2\pi}\}$, so letting $N \to \infty$ gives $c_n = 0$ for all $n \neq 0$. (The uniformity of convergence allows the interchange of limit and integral.) Q.E.D.

Cantor's beautiful theorem suggests a variety of extensions and generalizations. The remaining four sections of this paper will consider some of these.

**2. Summability and uniqueness.** Can we improve Theorem C by weakening the assumption of convergence to zero to an assumption of being merely summable to zero? As is often the case in mathematics, the starting point is a counterexample which destroys the "obvious extension."

The trigonometric series $\Sigma c_n e^{inx}$ is said to be Abel summable to $s$ if for each $r$, $0 \leqslant r < 1$, $f(x, r) := \Sigma c_n e^{inx} r^{|n|}$ converges and if $\lim_{r \to 1^-} f(x, r) = s$. Let $z := re^{ix} = r(\cos x + i \sin x)$. Differentiate the identity

$$\sum_{n=0}^{\infty} (\cos nx) r^n = \mathcal{R}e\left\{\sum_{n=0}^{\infty} z^n\right\} = \mathcal{R}e\left\{\frac{1}{1-z} \cdot \frac{1-\bar{z}}{1-\bar{z}}\right\} = \frac{1 - r\cos x}{1 - 2r\cos x + r^2}$$

with respect to $x$ to obtain

$$\sum_{n=1}^{\infty} (n \sin nx) r^n = \frac{(1-r^2) r \sin x}{(1 - 2r\cos x + r^2)^2}.$$

If $x \neq 0$, as $r \to 1^-$, the right side tends to 0; while if $x = 0$, every term of $f(0, r)$ is 0, so that $f(0, r) = 0$, whence $\lim_{r \to 1^-} f(0, r) = 0$. Thus $\Sigma_{n=1}^{\infty} n \sin nx$ is *everywhere* Abel summable to 0. Although this example is unpleasant, it turns out to be just about the worst thing that can happen.

THEOREM V (VERBLUNSKY [25, VOL. I, PP. 352, 383], [22], [23]). *If $c_n/|n| \to 0$ as $|n| \to \infty$ and $\Sigma c_n e^{inx}$ is Abel summable to 0 at every $x$, then all $c_n$ are 0.*

**3. Higher dimensions.** When we move from one to several dimensions, the picture becomes much more cloudy. Here is the land of opportunity. Almost nothing is known; almost every question that the novice might ask turns out to be an open question. The first goal is to mimic Cantor's Theorem C. Even this apparently modest goal remains to a large extent unachieved. The hypothesis that $\Sigma c_n e^{inx}$ converges to 0 has several different interpretations in higher dimensions. Most of these can be illustrated in two dimensions, so to ease notation I will restrict myself to that case. The basic object will be the double trigonometric series $T(x, y) := \Sigma_{(m,n) \in \mathbf{Z} \times \mathbf{Z}} c_{mn} e^{i(mx+ny)}$. We define a rectangular partial sum of $T$ to be

$$T_{mn}(x, y) := \sum_{|\mu| \leqslant m,\, |\nu| \leqslant n} c_{\mu\nu} e^{i(\mu x + \nu y)},$$

a diamond shaped partial sum to be

$$T^n(x, y) := \sum_{|\mu| + |\nu| \leqslant n} c_{\mu\nu} e^{i(\mu x + \nu y)},$$

and a circular partial sum to be

$$T_r(x, y) := \sum_{\mu^2+\nu^2 \leqslant r^2} c_{\mu\nu} e^{i(\mu x+\nu y)}.$$

We freeze $x$ and $y$ and make five different definitions of convergence.
If $T_r \to s$ as $r \to \infty$ say $T$ is *circularly convergent* to $s$.
If $T^n \to s$ as $r \to \infty$ say $T$ is *triangularly convergent* to $s$.
If $T_{nn} \to s$ as $r \to \infty$ say $T$ is *square convergent* to $s$.
If $T_{mn} \to s$ as $\min\{m, n\} \to \infty$ say $T$ is *unrestrictedly rectangularly convergent* to $s$.
If $T_{mn} \to s$ as $\min\{m, n\} \to \infty$ in such a way that $m/n$ and $n/m$ stay less than $e$, and if this happens for each (arbitrarily large) $e > 1$, say $T$ is *restrictedly rectangularly convergent* to $s$.

To each of these five notions of convergence there corresponds a putative extension of Theorem C. The first of these to be proved was the following.

THEOREM SC (V. Shapiro and R. Cooke **[19]**, **[7]**). *If $T(x, y)$ is circularly convergent to 0 everywhere, then all $c_{mn}$ are 0.*

In 1957 Victor Shapiro proved a two-dimensional version of Theorem V which implied a weak version of Theorem SC that required the additional hypothesis that

$$\frac{1}{r} \sum_{(r-1)^2 < m^2+n^2 \leqslant r^2} |c_{mn}| \to 0 \quad \text{as} \quad r \to \infty. \tag{1}$$

That this hypothesis was not needed was a consequence of a generalization of the Cantor-Lebesgue theorem due to Roger Cooke in 1971. (See Theorem A1 below. A survey of Cooke's theorem and extensions of it by Zygmund **[26]** and Connes **[6]** can be found in the MONTHLY article by Cooke **[8]**.)

The proof of Theorem SC is modeled after Verblunsky's Theorem V mentioned above. Both these theorems carry out the Riemann-Cantor program of integrating twice and then differentiating twice. To get at the ideas behind Theorem SC, assume that $T$ is circularly convergent to zero everywhere. Write $M := (m, n)$, $X := (x, y)$, $M \cdot X = mx + ny$, and add the assumption that $c_{00} = 0$. (This simplifies the notation, but not the proof.)

Let

$$F(X, t) := -\sum_{M \neq 0} \frac{c_M}{|M|^2} e^{iM\cdot X - |M|t}$$

and let

$$F(X) := \lim_{t \to 0} F(X, t).$$

Then *formally*

$$F(X)\text{``}=\text{''} -\sum_{M \neq 0} \frac{c_M}{|M|^2} e^{iM\cdot X} =: S$$

and *formally* $F(X)$ is a second integral of $T(X)$.

From Cooke's two-dimensional Cantor-Lebesgue theorem it follows that equation (1) holds. This guarantees the existence of $F(X, t)$. That $F(X)$ exists for all $X$ follows from the convergence and consequent circular Abel summability of $T$ **[18**,

pp. 67–68]. We have

$$\sum_{M\neq 0}\frac{|c_M|^2}{|M|^4}\leqslant \sum_{|M|=1}|c_M|^2+\sum_{k=2}^{\infty}\frac{1}{(k-1)^4}\left\{\sum_{k-1<|M|\leqslant k}|c_M|^2\right\}$$

Condition (1) implies $|c_M| = \circ(|M|)$ so applying first this and then (1) itself shows the curly bracketed sum to be $\circ(k^2)$. Hence the sum is finite, so that by the Riesz-Fischer Theorem [**20**, p. 248] $S$ is the Fourier series of a square integrable function. A basic fact about circular Abel summability implies that the function in question must be $F$ [**20**, Cor. 2.15, p. 256]. In short $\{c_M/|M|^2\}$ are the Fourier coefficients of the square integrable function $F(X)$.

The first major obstacle is that in distinction to the one dimensional case, it is quite difficult to demonstrate the continuity of $F$.

In fact if $F$ is continuous on the closure of an open disc $B$, then $F$ is actually harmonic on $B$. This depends on an argument very much like the one given in appendix 2. The substitute for the second Schwarz derivative $D$ used there is the generalized Laplacian $\Delta$ defined by

$$\Delta F(X) := \lim_{h\to 0}\frac{8}{h^2}\{F_h(X)-F(X)\},$$

where

$$F_h(X) := \frac{1}{\pi h^2}\int_{|H|\leqslant h}F(X+H)\,dH.$$

This agrees with the usual Laplacian ($=\partial^2F/\partial x^2+\partial^2F/\partial y^2$) for $C^2$ functions as can be seen by expanding $F$ into its Taylor series. We also have the representation

$$F(X,t)=\frac{t}{2\pi}\int_{R^2}\frac{F(U)}{\left[t^2+(X-U)^2\right]^{3/2}}\,dU \quad [\mathbf{18}, \text{p. } 56].$$

Changing to polar coordinates and integrating by parts gives

$$F(X,t)=\frac{3}{2}t\int_0^{\infty}\frac{r^3F(X)}{(t^2+r^2)^{5/2}}\,dr \quad [\mathbf{18}, \text{pp. } 66\text{–}67]. \tag{2}$$

Now

$$\frac{d^2}{dt^2}[F(x,t)] = -\sum c_M e^{iM\cdot X-|M|t}$$

is the circular Abel means of the original series, so that from the hypothesis (1) it is immediate that $\lim_{t\to 0}(d^2/dt^2)[F(x,t)]=0$. Shapiro then adapts a clever one-dimensional lemma of Rajchman [**25**, vol 1, pp. 353–354] to conclude from this that $\Delta F(X)=0$ at each $X$. This implication depends heavily on equation (2) [**18**, pp. 66–67].

By an argument like the one in appendix 2, zero generalized Laplacian forces harmonicity. (The analog of the subtracted linear function of appendix 2 is the Poisson integral of $F$.) See Radó for details [**12**, p. 14]. It follows that if $F(X)$ can be shown continuous everywhere, it will be harmonic everywhere. The continuous function $F$ will be necessarily bounded on the compact set $[0,2\pi]\times[0,2\pi]$ and

consequently by periodicity bounded on the entire plane. But bounded harmonic functions are constant. (Apply Liouville's Theorem to the bounded analytic function $\exp(F + i\tilde{F})$, where $\tilde{F}$ is a harmonic conjugate of $F$.) So there will be a constant $d$ so that

$$d + \sum_{M \neq 0} \frac{c_M}{|M|^2} e^{iM \cdot X} = 0$$

for all $X$. By the uniqueness theorem for square integrable functions (an immediate consequence of Parseval's formula [**25**, vol. II, p. 301]) it will follow that all $c_M = 0$.

It remains only to establish the continuity of $F$. This is done by generalizing a Baire category argument employed by Verblunsky.

Since $T(x, t)$ the circular Abel mean of the original series is continuous on $\{(X, t)\colon t > 0\}$ and has finite limit (namely 0) at each point, by a basic theorem of Baire for each disc in $T^2$ there is a subdisc and a constant $K$ so that $|T(X, t)| \leqslant K$ for every $X$ of the subdisc and every positive $t$ [**25**, Vol. I, p. 29 (12.3i)] and [**18**, pp. 69–70]. Integration in the $t$ variable then shows that in that subdisc $F(X)$ is continuous, being the uniform limit of $F(X, t)$.

Let $Z$ be the set where $F$ is not continuous. Let $\bar{Z}$ be the closure of $Z$. Assume $Z$ is non-empty. An inspired idea that appears already in the proof of Theorem V now produces the desired contradiction, as follows. Applying the same Baire category argument that showed $F(X, t)$ well-behaved on an open dense set to the set $\bar{Z}$ produces a point $X_0 \in Z$ with $F(X, t)$ converging uniformly to $F(X)$ with respect to $\bar{Z}$ throughout a neighborhood of $X_0$. If $X$ is any point very close to $X_0$, let $X_1$ be a point of $\bar{Z}$ closest to $X$, say $|X_1 - X| = s$. Assume $X \neq X_1$, $X_1 \neq X_0$. (If either equality holds, the argument is even easier.) Then (i) $F(X) = F_s(X)$, (ii) $F_s(X)$ is close to $F_s(X_1)$, (iii) $F_s(X_1)$ is close to $F(X_1, s)$, (iv) $F(X_1, s)$ is close to $F(X_1)$, and (v) $F(X_1)$ is close to $F(X_0)$; whence $F$ is continuous at $X_0$, contrary to the definition of $Z$. (Reasons: (i) Continuity forces harmonicity as mentioned above, and harmonicity is equivalent to the mean value property ([**12**], p. 7). (ii), (iii) These need technical lemmas whose proofs are straightforward provided one is aware of the delicate estimate

$$\left| \int_{|X| \leqslant 1} e^{iX \cdot t}\, dX \right| \leqslant \frac{C}{|t|^{3/2}}$$

for large $t$. This is equivalent to the fact that the Bessel function $J_1(s) = 0(s^{-1/2})$ as $s \to +\infty$ [**18**, pp. 68–69] and [**20**, p. 199]. (iv) Uniformity of convergence gives this. (v) Uniform limits are continuous.)

This completes our discussion of the proof of Theorem SC.

The other major result in two dimensions is this.

THEOREM AW (J. M. Ash and G. Welland) [**1**], [**2**]. *If $T(x, y)$ is unrestrictedly rectangularly convergent to* 0 *everywhere, then all $c_{mn}$ are* 0.

From the hypothesis it is immediate that at each $X$ the partial sums tend to 0 "in the northeast," i.e., that

$$\lim_{\min\{m, n\} \to \infty} T_{mn}(X) = 0.$$

One-dimensional convergent sequences of numbers are necessarily bounded, but two dimensional ones need not be. Nevertheless it can be proved that at each $X$,

$\sup_{m,n}|T_{mn}(X)| < \infty$. In fact this is the hardest part of the proof of Theorem AW. It requires a technique that first appeared in the unpublished thesis of P. J. Cohen [**2**, pp. 402, 404–407].

Write

$$C_{mn}(X) := C_{m,n}e^{i(mx+ny)} + C_{-m,n}e^{i(-mx+ny)} + C_{m,-n}e^{i(mx-ny)} + C_{-m,-n}e^{i(-mx-ny)}.$$

The "Mondrian" identity

$$C_{mn}(X) = T_{mn}(X) - T_{m-1,n}(X) - T_{m,n-1}(X) + T_{m-1,n-1}(X)$$

implies that at each $X$ the $C_{mn}(X)$ also tend to 0 in the northeast and are bounded [**2**, pp. 410–411]. A Cantor-Lebesgue theorem finally gets back to the coefficient themselves. The result is that $\{c_{mn}\}$ is a bounded sequence and $\lim_{\min\{|m|,|n|\}\to\infty} c_{mn} = 0$ [**2**, p. 408 and p. 411]. These two facts allow one to readily deduce that hypothesis (1) holds [**2**, p. 423]. It is also true that a unrestrictly rectangularly convergent double sequence of numbers is circularly Abel summable to the same value, *provided the partial sums are bounded* [**2**, pp. 413–416]. In particular, since $T(X)$ converges unrestrictly rectangularly to 0 everywhere, it is circularly Abel summable to 0 everywhere. A careful examination of the proof of Theorem SC shows that condition (1) together with everywhere circular Abel summability to 0 are sufficient hypotheses for the proof to work.

*Remarks.* When Grant Welland and I discovered this proof, I felt that there was an element of good luck involved here. First, the rectangular convergence gave just enough control of the coefficient size to force condition (1) and condition (1) is, in a sense, sharp. (The series $T_0(x, y) := \Sigma n \sin nx$ "almost" satisfies (1) and is circularly Abel summable to 0 everywhere. The computations are in section 2 above.) More impcrtantly, there are usually no nontrivial connections between various modes of convergence and in particular unrestricted rectangular convergence of a trigonometric series on a set does not force circular convergence of the series, even if one is willing to discard a subset of measure 0 [**2**, pp. 417–420]. Thus the above mentioned result connecting unrestricted rectangular convergence and circular Abel summability came as a happy surprise.

The other side of the coin is that this proof of Theorem AW was "too lucky." Often rectangular results for two-dimensional series can be extended to higher dimensions without much additional effort. The unexpected dependence of Theorem AW on Theorem SC means that a good three-dimensional uniqueness theorem for unrestricted rectangular convergence awaits either a good three-dimensional spherical uniqueness theorem or a completely new method of proof. Shapiro has proved a three-dimensional spherical uniqueness theorem, but it needs the hypothesis

$$\frac{1}{r^\alpha} \sum_{(r-1)^2 < l^2+m^2+n^2 \leqslant r^2} |c_{lmn}| \to 0 \quad \text{as} \quad r \to \infty \qquad (N_\alpha)$$

with $\alpha = 1$. But a three-dimensional Cantor-Lebesgue theorem can only be expected to produce condition $(N_\alpha)$ with $\alpha = 2$. (See [**2**, p. 425] for a relevant counterexample.) Thus even if one assumes that $T(x, y, z)$ is everywhere circularly convergent to 0, it is not known whether all $c_{lmn}$ must then be 0. There is even greater ignorance

concerning uniqueness questions about the other 3 modes of convergence mentioned above. For example, here are 3 open questions in two dimensions.

Question 1. If $T(x, y)$ is everywhere restrictedly rectangularly convergent to 0, does this force all $c_{mn}$ to be 0?
Question 2. If $T(x, y)$ is everywhere square convergent to 0, does this force all $c_{mn}$ to be 0?
Question 3. If $T(x, y)$ is everywhere triangularly convergent to 0, does this force all $c_{mn}$ to be 0?

Again, it must be emphasized that easy counterexamples show that not much help will be available from Cantor-Lebesgue type theorems for square, restricted rectangular, or triangular convergence [**2**, pp. 416–418].

**4. Fourier series.** Return to one dimension. A third type of extension of Cantor's Theorem C occurs when the limit of the trigonometric series $S := \Sigma c_n e^{inx}$ is a Lebesgue integrable function $f$. The question now is whether $S$ is necessarily the Fourier series of $f$, i.e., whether for each integer $n$ there must hold the relation

$$c_n = \frac{1}{2\pi}\int_0^{2\pi} f(x)e^{-inx}\,dx.$$

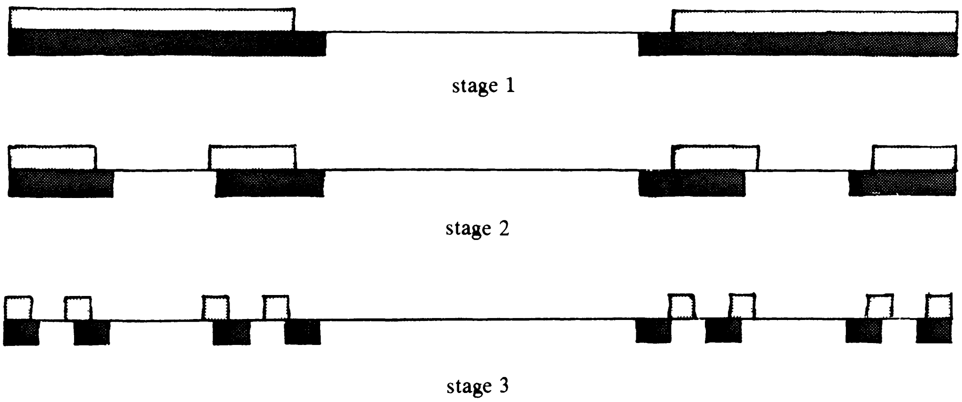
A typical result in this direction is the following.

THEOREM (de la Vallée-Poussin [**25**, Vol. I, pp. 326, 382], [**21**]). *If $S$ converges to $f$ at each $x$, and if $f$ is finite at each $x$ and if $\int_0^{2\pi}|f(x)|\,dx < \infty$, then $S$ is the Fourier series of $f$.*

**5. Sets of uniqueness.** If we replace the hypothesis $\Sigma c_n e^{inx} = 0$ everywhere by $\Sigma c_n e^{inx} = 0$ for $x$ *not* in a "thin" subset $E$ of $[0, 2\pi]$, we may still be able to conclude that all $c_n = 0$ if $E$ is thin enough. Sufficiently thin sets $E$ are called sets of uniqueness. A restatement of Theorem C is that the empty set is a set of uniqueness. It was already proved by W. H. Young [**24**] that every countable subset of $[0, 2\pi]$ is a set of uniqueness. (Cantor's earlier work showing that closed countable sets were sets of uniqueness led Cantor to the creation of set theory!) However, a set $E$ of positive measure is too thick to be a set of uniqueness. (Proof from [**25**, Vol. I, p. 344]: Let $E_1$ be a subset of $E$ which is perfect and of positive measure, and let $f(x)$ be the characteristic function of $E_1$. The Fourier series of $f$ converges to 0 outside $E_1$, and so also outside $E$, but does not vanish identically since its constant term is $|E_1|/2\pi > 0$.) One of the most interesting theorems in all analysis produces two classes of uncountable measure zero sets which are very much like each other metrically, although the first are sets of uniqueness and the second are not.

Let $0 < \xi < 1/2$. Dissect $[0, 2\pi]$ into 2 closed "white" intervals, each of length $2\pi\xi$, $[0, 2\pi\xi]$ and $[2\pi(1-\xi), 2\pi]$, and one open "black" interval $(2\pi\xi, 2\pi(1-\xi))$. Remove the black interval and repeat the process by dissecting each white interval into 2 closed white intervals of length $2\pi\xi^2$, and 1 centered open black interval of length $2\pi\xi - 2\cdot 2\pi\xi^2$. Iterating this process $k$ times produces $2^k$ closed white intervals, each of length $2\pi\xi^k$. (See FIGURE 1.)

Now let $k \to \infty$ remembering to remove the black intervals at each stage. The resulting set $E(\xi)$ is said to be of Cantor type with constant ratio of dissection.




□ "white" intervals for $E(1/3)$
■ "white" intervals for $E(1/3)$

FIG. 1.

Since $2^k \cdot 2\pi\xi^k \to 0$ as $k \to \infty$, $E(\xi)$ always has measure zero. In particular $E(1/3)$ is the classical Cantor set.

THEOREM SZ (R. Salem and A. Zygmund **[16]**, **[17]**, **[25**, Vol. II, p. 152], **[15]**, **[11]**). *If $1/\xi$ is an algebraic integer all of whose conjugates have modulus less than* 1 (*i.e.*, $(1/\xi)^n + a_1(1/\xi)^{n-1} + \cdots + a_n = 0$ *for some integers* $a_1, a_2, \ldots, a_n$, *$n$ is minimal, and the other $n - 1$ roots of $x^n + a_1x^{n-1} + \cdots + a_n = 0$ have absolute value less than* 1), *then $E(\xi)$ is a set of uniqueness. Otherwise $E(\xi)$ is not a set of uniqueness.*

Now $1/3 = .33\ldots > .3 = 3/10$ so the process that forms $E(1/3)$ leaves more material at each stage than does the process that forms $E(3/10)$. Hence $E(1/3)$ is intuitively thicker than $E(3/10)$. Nevertheless $1/(1/3)$ satisfies $x - 3 = 0$ and hence is an algebraic integer without conjugates so that $E(1/3)$ is a set of uniqueness; while $1/(3/10)$ is not an algebraic integer at all[3] so that $E(3/10)$ is not a set of uniqueness. Thus the "wrong" one of the pair $E(1/3)$, $E(3/10)$ is the set of uniqueness.

Even more dramatically counterintuitive is the fact that the "very thick" set $\bigcup E(\xi)$, where $\xi$ varies over all reciprocals of algebraic integers with small conjugates ($\xi = 1/2$ must be excluded), is also a set of uniqueness. This is true because N. Bary has proved that a countable union of closed sets of uniqueness is still a set of uniqueness. See (**[25]**, Vol. I, p. 349) for this fact as well as further remarks on Theorem SZ.

The other side of the coin from a set of uniqueness is a set of multiplicity. A set of multiplicity by definition is a subset of $[0, 2\pi]$ which is not a set of uniqueness. More directly, $E$ is a set of multiplicity if there is a trigonometric series which converges to 0 outside $E$ but which does not vanish identically. Just as sets of

[3]Assume $10/3$ satisfies $x^n + \cdots + a_n = 0$ with all $a_i$ integer. Clearly $n \geqslant 2$, so $(10^{n-1} + \cdots + a_{n-1}3^{n-1})10 + a_n 3^n = 0$. Hence 3 divides $10^{n-1} + 3(a_1 + \cdots + a_{n-1}3^{n-2})$, so 3 divides $10^{n-1}$, a contradiction.

uniqueness become more interesting as they get "thicker", so sets of multiplicity become more interesting as they get "thinner." As was shown above, every set of positive measure is a set of multiplicity. The first measure zero set of multiplicity was produced by Men'shov [**10**] and Theorem SZ above gives lots of examples.

In a later review I will survey some of the extensive recent work that has been done on sets of uniqueness. At present this is the most active and exciting area of the four extensions I have discussed. The interested reader is encouraged to turn now to the books of Zygmund [**25**] or Bary [**3**] for much more comprehensive overviews of all the one-dimensional topics I have highlighted here.

*Note added in proof.* A remarkable facet of Cantor's proof of Theorem C has been *its* uniqueness. The many difficulties encountered in attempts to generalize it suggest that a different proof of Theorem C could prove useful. Because of a recent development in real analysis I can now give such a proof. Suppose the series $\Sigma c_n e^{inx}$ converges to zero everywhere. Form the *first* integral $c_0 x + \Sigma(c_n/in)e^{inx}$. Although this $L^2$ function is not easily seen to be continuous (The examples $\Sigma \sin n0/\ln n = 0$ but $-\Sigma \cos n0/n \ln n$ divergent give the flavor of the difficulty in working with a first integral.), it does follow directly from a theorem of Rajchman and Zygmund [**25**, Vol. 1, p. 324] that it has symmetric approximate derivative 0 at every point. By the aforementioned recent results {"A symmetric density property; monotonicity and the approximate symmetric derivative," *Proc. Amer. Math. Soc.*, 104 (1988) 1078–1102, and "A symmetric density property for measurable sets," *Real Analysis Exchange*, 14 (1988–89) 203–209, both by C. Freiling and D. Rinne}, there is a constant $c$ so that

$$c_0 x + \sum \frac{c_n}{in} e^{inx} - c = 0$$

almost everywhere. The conclusion of Theorem C now follows from periodicity and Plancherel's Theorem. I will publish the details of this proof in the *Proc. Amer. Math. Soc.*

*Appendix* 1. THEOREM A1 (Cantor [**4**]). *If* $c_n e^{inx} + c_{-n} e^{-inx} \to 0$ *as* $n \to \infty$ *for every* $x$, *then* $|c_n| + |c_{-n}| \to 0$.

*Proof.* From the convergence to zero of $c_n e^{inx} + c_{-n}e^{-inx}$ follows the convergence to zero of its real and imaginary parts. One may easily find real numbers $a_n, b_n, a'_n, b'_n$, so that

$$c_n e^{inx} + c_{-n} e^{-inx} = (a_n \cos nx + b_n \sin nx) + (a'_n \cos nx + b'_n \sin nx)i$$

and direct calculation shows

$$a_n^2 + b_n^2 + a_n'^2 + b_n'^2 = 2(|c_n|^2 + |c_{-n}|^2).$$

Hence it suffices to prove that $a_n \cos nx + b_n \sin nx \to 0$ for every $x$ implies $a_n^2 + b_n^2 \to 0$. Define $\rho_n := \sqrt{a_n^2 + b_n^2}$ and find $\theta_n$ so $\rho_n \cos\theta_n = a_n$, $\rho_n \sin\theta_n = b_n$. Then $a_n \cos nx + b_n \sin nx = \rho_n \cos(nx - \theta_n)$, $\rho_n \cos(nx - \theta_n) \to 0$ for every $x$, and we need only show that $\rho_n \to 0$.

If $\rho_n$ does not tend to 0, there is a subsequence $\{n_k\}$ and a number $\delta$ so that for every positive integer $k$, $\rho_{n_k} \geqslant \delta > 0$.

By discarding as many terms as necessary from $\{n_k\}$ we may also assume $(n_{k+1}/n_k) \geqslant 3$ for all $k$. Then

$$\cos(n_1 x - \theta_{n_1}) \geqslant \frac{1}{2} \qquad \text{for } x \in I_1 := \left[\left(\theta_{n_1} - \frac{\pi}{3}\right)\Big/ n_1, \left(\theta_{n_1} + \frac{\pi}{3}\right)\Big/ n_1\right].^4$$

Since

$$|I_1| = \frac{2\pi}{3n_1} \quad \text{and} \quad \frac{n_2}{n_1} \geqslant 3,$$

as $x$ ranges over $I_1$, $(n_2 x - \theta_{n_2})$ ranges over $n_2 I_1 - \theta_{n_2}$ and

$$|n_2 I_1 - \theta_{n_2}| = n_2 |I_1| = n_2 \cdot \frac{2\pi}{3n_1} \geqslant 2\pi.$$

Thus there is a closed interval $I_2 \subset I_1$ with $\cos(n_2 x - \theta_2) \geqslant 1/2$ for all $x \in I_2$ and $|I_2| = 2\pi/3n_2$. Proceeding inductively produces a point $\xi \in \bigcap_{k=1}^{\infty} I_k$ with $\cos(n_k \xi - \theta_{n_k}) \geqslant 1/2$ for every $k$. Thus $\rho_{n_k} \cos(n_k \xi - \theta_{n_k}) \geqslant \delta/2$ for every $k$. This contradicts the convergence to zero of $\rho_n \cos(nx - \theta_n)$ at $x = \xi$.

*Appendix* 2. THEOREM A2 (Schwarz, Cantor **[5]**, **[9**, p. 33]. *A continuous function $G$ with everywhere* 0 *Schwarz derivative is necessarily a linear function.*

*Proof.* First suppose that there is a continuous nonconvex function $H(x)$ satisfying $DH(x) > 0$. Then there are points $a < b < d$ with $H(b) > L(b)$, where $L$ is the linear function whose graph passes through $(a, H(a))$ and $(d, H(d))$. Let $H_1(x) := H(x) - L(x)$. Then $H_1(a) = H_1(d) = 0$, $H_1(b) > 0$, $H_1$ is continuous and $DH_1(x) = DH(x) > 0$ for all $x$. Let $c$ be a point of $[a, d]$ where $H_1$ is maximum. Note $c \in (a, d)$. (See Figure 2.) Then for all $h \leqslant \min\{c - a, d - c\}$, $(1/2)[H_1(c + h) + H_1(c - h)] - H_1(c) \leqslant 0$, contrary to $DH_1(c) > 0$.

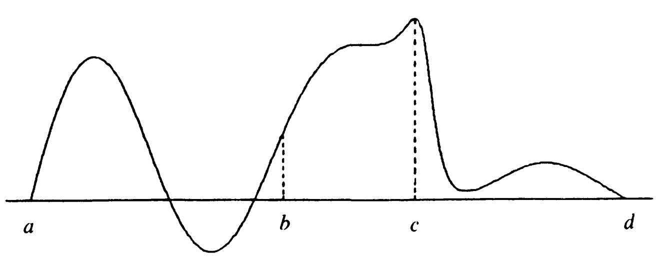



FIG. 2.

Now from $D(x^2) = 2$ (an easy calculation) and $DG = 0$, for each $\varepsilon > 0$ we have $D(G + \varepsilon x^2)(x) > 0$ for all $x$, so $G + \varepsilon x^2$ is convex. Let $\varepsilon \to 0$ to see that $G$ is convex. Symmetrically, $G - \varepsilon x^2$ is concave and hence $G$ is also. Since the graph of

[4] Warning: Here we must think of $[0, 2\pi]$ as $R/2\pi Z$, i.e. as the circumference of the unit circle. All this really means is that the set $I_1$ should be thought of as an interval of length $2\pi/3n_1$, even if $\theta_{n_1}/n_1$ is less than $\pi/3n_1$ from 0 or $2\pi$.

$G$ lies both above and below the chord passing through any two of its points, it is a line.

*Appendix* 3. THEOREM A3 (Riemann [**13**, §8, Theorem 1]. *Let*

$$s_{nk} := \left(\frac{\sin nk}{nk}\right)^2, \qquad n > 0, \qquad s_{0k} := 1$$

*and suppose* $\sum_{n=0}^{\infty} a_n = 0$. *Let* $G_k = \sum_{n=0}^{\infty} a_n s_{nk}$. *Then* $\lim_{k\to\infty} G_k = 0$.

*Proof.* For each fixed $k \neq 0$, the series defining $G_k$ converges since $a_n \to 0$ and $|s_{nk}| \leqslant c/n^2$. Summation by parts yields

$$\sum_{n=0}^{N} a_n s_{nk} = \sum_{n=0}^{N-1} s_n(s_{nk} - s_{n+1k}) + s_N s_{Nk},$$

where $s_n = a_0 + \cdots + a_n$. Let $N \to \infty$ to get

$$G_k = \sum_{n=0}^{\infty} s_n(s_{nk} - s_{n+1k}).$$

Now for each integer $N \geqslant 1$ and each $k \neq 0$,

$$|G_k| \leqslant \sup|s_n| \cdot \sum_{n=0}^{N-1} (|s_{nk} - 1| + |1 - s_{n+1k}|) + \left(\sup_{n\geqslant N}|s_n|\right) \sum_{n=N}^{\infty} |s_{nk} - s_{n+1k}|$$
$$=: A + B.$$

Since $|s_{nk} - s_{n+1k}| = |\int_{nk}^{(n+1)k} f(x)\,dx|$ where $f(x) := \{[\sin x/x]^2\}'$, the sum in $B$ is bounded by $\int_0^{\infty} |f(x)|\,dx$. This is a finite constant since

$$f(x) = 2\left(\frac{\sin x}{x}\right)\left(\frac{x\cos x - \sin x}{x^2}\right),$$

$$x\cos x - \sin x \cong x\left(1 - \frac{x^2}{2}\right) - \left(x - \frac{x^3}{6}\right) = -\frac{x^3}{3}$$

near $x = 0$ so that $f$ is bounded on $(0,1]$, and $|f(x)| \leqslant 2 \cdot 1 \cdot (x \cdot 1 + 1)/x^3 \leqslant 4/x^2$ for $x \geqslant 1$. It follows that $B$ can be made small by choosing $N$ large. Once $N$ is fixed, $A$ can be then made small by picking $|k|$ small.

REFERENCES

1. J. M. Ash and G. V. Welland, Convergence, summability, and uniqueness of multiple trigonometric series, *Bull. Amer. Math. Soc.*, 77 (1971) 123–127. MR 43#812.
2. ______, Convergence, uniqueness, and summability of multiple trigonometric series, *Trans. Amer. Math. Soc.*, 163 (1972) 401–436. MR 45#9057.
3. N. K. Bary, A Treatise on Trigonometric Series, Vol. 2, MacMillan, New York, 1964. MR 30#1347.
4. G. Cantor, Über einen die trigonometrischen Reihen betreffenden Lehrsatz, *Crelles J. für Math.*, 72 (1870) 130–138; also in Gesammelte Abhandlungen, Georg Olms, Hildesheim, 1962, pp. 71–79.
5. ______, Beweis, das eine für jeden reellen Wert von $x$ durch eine trigonometrische Reihe gegebene Funktion $f(x)$ sich nur auf eine einzige Weise in dieser Form darstellen lässt, *Crelles J. für Math.*, 72 (1870) 139–142; also in Gesammelte Abhandlungen, Georg Olms, Hildesheim, 1962, pp. 80–83.
6. B. Connes, Sur les coefficients des séries trigonométriques convergentes sphériquement, *C. R. Acad. Sci. Paris*, Ser A, 283 (1976) 159–161. MR 54#10975.

7. R. Cooke, A Cantor-Lebesgue theorem in two dimensions, *Proc. Amer. Math. Soc.*, 30 (1971), 547–550. MR 43#7847.
8. ______, The Cantor-Lebesgue theorem, this MONTHLY, 86 (1979) 558–565. MR 81b:42019.
9. J. W. Dauben, Georg Cantor: His Mathematics and Philosophy of the Infinite, Harvard Univ. Press, Cambridge, Mass., 1979. MR 80g:01021.
10. D. Men'shov, Sur la convergence uniforme des séries de Fourier (in Russian), *Mat. Sbornik*, 11 (1942) 67–96. MR 3-106.
11. Y. Meyer, Algebraic Numbers and Harmonic Analysis, North-Holland, Amsterdam, and American Elsevier, New York, 1972. MR 58#5579.
12. T. Radó, Subharmonic Functions, Chelsea, New York, 1949.
13. B. Riemann, Über die Darstellbarkeit einer Function durch eine trigonometrische Reihe, Gesammelte Abhandlungen, 2. Aufl., Georg Olms, Hildesheim, 1962, pp. 227–271; also Dover, New York, 1953.
14. K. A. Ross, Elementary Analysis: The Theory of Calculus, Springer-Verlag, New York, 1980. MR 81a:26001.
15. R. Salem, Algebraic Numbers and Fourier Analysis, D. C. Heath, Boston, 1963. MR 28#1169.
16. R. Salem and A. Zygmund, Sur un théorème de Piatetski-Shapiro, *C. R. Acad. Paris*, 240 (1955) 2040–2042. MR 17-150.
17. ______, Sur les ensembles parfaits dissymétriques à rapport constant, ibid *C. R. Acad. Paris*, 240 (1955) 2281–2283. MR 17-150.
18. V. L. Shapiro, Fourier series in several variables, *Bull. Amer. Math. Soc.*, 70 (1964) 48–93. MR 28#1448.
19. ______, Uniqueness of multiple trigonometric series, *Ann. of Math.*, (2) 66 (1957) 467–480. MR 19-854, 1432.
20. E. M. Stein and G. Weiss, Introduction to Fourier Analysis on Euclidean Spaces, Princeton Univ. Press, Princeton, 1971. MR 46#4102.
21. Ch. J. de la Vallée-Poussin, Sur l'unicité du développement trigonométrique, *Bull. Acad. Roy. de Belg.* (1912) 702–718.
22. S. Verblunsky, On the theory of trigonometric series, I, *Proc. Lon. Math. Soc.*, 34 (1932) 441–456.
23. ______, On the theory of trigonometric series, II, *Proc. Lon. Math. Soc.*, 34 (1932) 457–491.
24. W. H. Young, A note on trigonometrical series, *Mess. for Maths.*, 38 (1909) 44–48.
25. A. Zygmund, Trigonometric series. Vols. 1, 2, 2nd rev. ed., Cambridge Univ. Press, New York, 1959. MR 21#6498.
26. ______, A Cantor-Lebesgue theorem for double trigonometric series, *Studia Math.*, 43 (1972) 171–175. MR 47#711.
27. ______, Notes on the history of Fourier series, Studies in Harmonic Analysis, Studies in Mathematics, Vol. 13, J. M. Ash, ed., Math. Assoc. of America, 1976, pp. 1–19. MR 56#12740.

# Optimal Inscribed Polygons in Convex Curves

JOHN S. LEW AND DONALD A. QUARLES, JR., *T. J. Watson Research Center, Yorktown Heights, NY 10598*

JOHN S. LEW received a B.S. (1955) in physics from Yale University, then a Ph.D. (1960) in physics from Princeton University. He spent two years in the U.S. Air Force doing research in New Mexico, held applied-mathematics teaching positions at MIT and Brown University, then, in 1970, joined the Mathematical Sciences Department at the IBM T. J. Watson Research Center. His research interests include integral transforms and asymptotic expansions; but the motivation for this paper was more recent work with Quarles to develop finite-element computer programs for magnetostatic problems.

DONALD A. QUARLES, JR. received his B.S. from Yale University in 1943, and returned there in 1946 for graduate work (M.A., 1948) and teaching after service in the U.S. Naval Reserve. He began work at IBM with a summer job in 1948, and later resumed graduate work during the evenings at New York University (Ph.D., 1964). He has worked in several areas, including numerical weather prediction, satellite tracking, shock wave calculations, and in the simulation of electromagnetic and electromechanical devices. His research activity was interrupted by a teaching sabbatical at Stevens Institute of Technology (1979–80).

**1. Problem Statement; Subdivision Algorithm.** For a given positive integer $n$, and a regular $C^2$ plane convex curve $\Gamma$ that includes no straight line segment, we wish to approximate $\Gamma$ by an inscribed $n$-segment polygonal line that minimizes the area between curve and polygonal line. For $n$ a power of 2, we show that Archimedes' approach to the quadrature of a parabola yields the correct result in any arc $\Gamma$ of a conic section. For arbitrary $n$, we give a slow but simple convergent iteration to a *locally* optimal solution. Indeed, this problem, unrestricted, may have multiple solutions. Hence we provide criteria that insure uniqueness, and we discuss a few examples. The motivation for this work was finding a graphical discretization of a plotted curve.

More specifically, $K$ in this article denotes a compact convex plane region whose boundary has two parts: (1) a straight-line segment $\Delta$, which may be trivial (i.e., a single point), and (2) a regular $C^2$ curve $\Gamma$ that includes no nontrivial straight-line segment. (Thus each tangent to $\Gamma$ intersects that curve at only a single point.) For the given integer $n$, we wish to inscribe in $\Gamma$ an $n$-segment broken line $\Gamma_n$ that, together with the given segment $\Delta$, forms a convex polygon with the largest possible area. Figure 1 shows a sample $\Gamma$, $\Delta$, and $\Gamma_n$. Earlier papers have considered the difference area or some other error measure, and shown convexity in the variable $n$ [Dowker 1944] or studied asymptotic behavior for large $n$ [Tóth 1948, McClure and Vitale 1975].

The problem statement needs no coordinate system. However, in the plane of the set $K$, we introduce standard $x$ and $y$ axes; and, perpendicular to this $xy$-plane, we add the usual third dimension with coordinate $z$. Throughout this section we take the origin strictly inside $K$. In the positive $x$, $y$, and $z$ directions we define the orthonormal vectors **i**, **j**, **k** of the standard notation. Clearly, the curve $\Gamma$ has a $C^2$

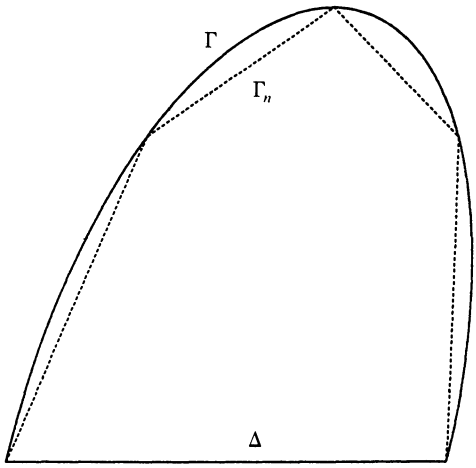



FIG. 1. Sample curve $\Gamma$ and segment $\Delta$ (solid) with inscribed polygon $\Gamma_n$ (dotted).

parametrization

$$\mathbf{r}(u) = x(u)\mathbf{i} + y(u)\mathbf{j}, \qquad a \leqslant u \leqslant b, \tag{1.1}$$

where $0 < b - a < +\infty$; indeed, the regularity assumption allows a parameter $u$ such that $d\mathbf{r}/du \neq 0$ and $\mathbf{r}(u)$ moves counterclockwise as $u$ increases. (For example, we may choose polar angle $\theta$ as the parameter.) Then curve $\Gamma$ and segment $\Delta$ both have endpoints $\mathbf{r}(a)$, $\mathbf{r}(b)$; and a trivial segment $\Delta$ would have $\mathbf{r}(a) = \mathbf{r}(b)$. Also, the curve $\Gamma$ has tangent vector $\mathbf{t}(u) = \mathbf{r}'(u)/|\mathbf{r}'(u)|$, where we define $(\cdot)' = (d/du)(\cdot)$; and this vector rotates counterclockwise as $u$ increases. Therefore, we can define the tangent angle $\tau(u)$ so that it is a strictly increasing $C^1$ function

$$\tau(u) = \arctan(y'(u)/x'(u)), \qquad \tau_0 \leqslant \tau(u) \leqslant \tau_0 + 2\pi. \tag{1.2}$$

Relevant geometric concepts have analytical equivalents. For any positive integer $n$, the sequence $(u_0, u_1, \dots, u_n)$ is a *partition* $\Pi_n$ of interval $[a, b]$ if

$$a = u_0 \leqslant u_1 \leqslant \cdots \leqslant u_{n-1} \leqslant u_n = b; \tag{1.3}$$

and $\Pi_n$ is *nondegenerate* if the $u_j$ are all distinct. We introduce the notation

$$\mathbf{r}_j = \mathbf{r}(u_j) \qquad \mathbf{r}'_j = \mathbf{r}'(u_j), \qquad \mathbf{r}''_j = \mathbf{r}''(u_j), \dots . \tag{1.4}$$

Then, in $\Gamma$, the point-sequence $\mathbf{r}_0\mathbf{r}_1 \dots \mathbf{r}_n$ specifies an arbitrary inscribed broken line $\Gamma_n$ with $n$ straight segments, which are all nontrivial precisely when $\Pi_n$ is nondegenerate. Also, the broken line $\mathbf{r}_0\mathbf{r}_1 \dots \mathbf{r}_n\mathbf{r}_0$ ($\Gamma_n$ with the additional segment $\Delta$) bounds a convex polygonal subset $K_n$ of $K$. If a radius vector from the origin traces curve $\Gamma$ (resp., $\Gamma_n$), then, by standard arguments (e.g., Green's Theorem), this vector sweeps out an area $F/2$ (resp., $F_n/2$), where

$$F = \mathbf{k} \cdot \int_a^b \mathbf{r}(u) \times \mathbf{r}'(u)\, du, \qquad F_n = \mathbf{k} \cdot \sum_{j=1}^{n} \mathbf{r}_{j-1} \times \mathbf{r}_j. \tag{1.5}$$

Thus $(F - F_n)/2 = \operatorname{area}(K \setminus K_n)$. We can now restate the problem: we wish to determine a (possibly degenerate) partition so as to maximize area$(K_n)$, hence to

minimize area$(K \setminus K_n)$, hence to maximize the area $F_n$. Here $K \setminus K_n$ denotes the set difference.

Clearly, $F_n(u_1, \ldots, u_{n-1})$ is a continuous function, and its domain is a compact set. Therefore, $F_n$ achieves a global maximum. However, $F_n$ may have two or more *local* maxima. (Section 4 gives two examples; Section 5 provides some uniqueness criteria.) Thus we call $\Pi_n$ a *locally optimal* partition if it yields a local maximum. Fortunately, a locally optimal $\Pi_n$ must be a *nondegenerate* partition, so that the corresponding $\mathbf{r}_0, \ldots, \mathbf{r}_n$ must be distinct vectors. Indeed, if $u_j$ equals either $u_{j-1}$ or $u_{j+1}$, while $u_{j-1} < u_{j+1}$, then, geometrically, any intermediate $u_j$ produces a larger area $(K_n)$—because $\Gamma$ includes no straight-line segments.

Equivalently, if $a \leqslant v < w \leqslant b$, then $\mathbf{r}'(u)$ passes *just once* through the direction of $\mathbf{r}(w) - \mathbf{r}(v)$ as $u$ increases from $v$ to $w$ (because $\tau(u)$ is a strictly increasing $C^1$ function). If we let $M(v, w)$ denote this unique intermediate value—where the tangent at $\mathbf{r}(M(v, w))$ parallels the chord between $\mathbf{r}(v)$ and $\mathbf{r}(w)$—then $v < M(v, w) < w$, and

$$0 = \mathbf{r}'(M(v, w)) \times [\mathbf{r}(w) - \mathbf{r}(v)]. \tag{1.6}$$

But (1.5) implies, directly,

$$\partial F_n / \partial u_j = \mathbf{k} \cdot \mathbf{r}'(u_j) \times [\mathbf{r}(u_{j+1}) - \mathbf{r}(u_{j-1})], \qquad j = 1, \ldots, n-1, \tag{1.7}$$

and $\tau(u)$ is strictly increasing; so if $u_{j-1} < u_{j+1}$, then $\partial F_n / \partial u_j > 0$ when $u_{j-1} \leqslant u_j < M(u_{j-1}, u_{j+1})$, and $\partial F_n / \partial u_j < 0$ when $M(u_{j-1}, u_{j+1}) < u_j \leqslant u_{j+1}$. Thus again a locally optimal $\Pi_n$ is a nondegenerate partition, and moreover any such $\Pi_n$ obeys the necessary conditions

$$u_j = M(u_{j-1}, u_{j+1}), \qquad j = 1, \ldots, n-1. \tag{1.8}$$

Section 3 presents a way to check local optimality, and describes a convergent iteration that reliably yields a locally optimal $\Pi_n$. However, conditions (1.8) suggest a construction that may offer a possible shortcut. If $\Pi_n$ is *any* nondegenerate partition, so that $a = u_0 < u_1 < \cdots < u_{n-1} < u_n = b$, then we define $u_{j-(1/2)} = M(u_{j-1}, u_j)$, where $j = 1, \ldots, n$, and, by doubling all subscripts, we obtain a *subdivided* partition $\Pi_{2n}$ such that $a = u_0 < u_1 < \cdots < u_{2n-1} < u_{2n} = b$. The corresponding broken lines $\Gamma_n$ and $\Gamma_{2n}$ have respectively $n$ and $2n$ straight segments; moreover, area$(K_{2n}) >$ area$(K_n)$. Hence we may take the coarsest possible partition $\Pi_1$, which has simply $a = u_0 < u_1 = b$, or we may take any other nondegenerate initial $\Pi_k$, and we may iterate this subdivision algorithm $m$ times to produce the finer nondegenerate partition:

$$a = u_0 < u_1 < \cdots < u_{n-1} < u_n = b, \qquad n = k \cdot 2^m. \tag{1.9}$$

By construction, the resulting *odd-numbered* $u_j$ satisfy the necessary conditions (1.8) for a locally optimal $\Pi_n$. Thus we ask the natural question: when do the *even-numbered* $u_j$ also satisfy these conditions, so that the subdivision algorithm yields a locally optimal partition?

This algorithm, hence this question, has both practical and historical interest. If one has merely a plot of the curve $\Gamma$, then one can readily construct the first few subdivisions within some visual accuracy. Having marked the points $\mathbf{r}_0, \mathbf{r}_1, \ldots, \mathbf{r}_n$ for the current $\Gamma_n$, lay a straightedge along each chord $\mathbf{r}_{j-1}\mathbf{r}_j$, and slide the straightedge parallel to itself until it appears to intersect $\Gamma$ at only a single tangency point

$\mathbf{r}_{j-(1/2)}$. Marking now all such tangency points gives approximately the sequence $(\mathbf{r}_0, \mathbf{r}_1, \dots, \mathbf{r}_{2n})$ that defines the subdivided $\Gamma_{2n}$. Heuristically, by this geometric method, the authors have chosen a plausible set of about 15 points to represent efficiently a plotted empirical curve in digital computations. *Computing* good locations for these points would have required numerical coordinate values for many additional points, and actually reading so much data from a mere plot would have been a tedious invitation to eyestrain.

Moreover, Archimedes used the subdivision concept about 2200 years ago to obtain some important geometric results. Clearly, subdivision, within a circle $\Gamma$, is just bisection of each arc $\mathbf{r}_{j-1}\mathbf{r}_j$. However, in a unit circle [Heath 1912, pp. 91–98], Archimedes inscribed a regular hexagon $\Gamma_6$, then repeatedly bisected its arcs to produce a regular 96-gon $\Gamma_{96}$, whose perimeter gave him a lower bound for $\pi$. As another application [Heath 1912, pp. 233–252], he considered the region $K$ between a parabola and a chord, took the chord itself as $\Gamma_1$, and by precisely the subdivision algorithm, constructed the successive polygonal lines $\Gamma_n$, where $n = 2^m$ and $m = 1, 2, \dots$. He then proved essentially that

$$\begin{aligned}\operatorname{area}(K) &= \lim\nolimits_n \operatorname{area}(K_n)\\ &= \lim\nolimits_m [1 + 4^{-1} + \cdots + 4^{1-m}]\operatorname{area}(K_2) = (4/3)\operatorname{area}(K_2), \quad (1.10)\end{aligned}$$

and thus he found area$(K)$, since $K_2$ is a triangle.

**2. Additive Parameters; Conic Sections.** Under an extra assumption we obtain first some results that answer most questions about the subdivision algorithm. An *additive* parameter for the curve $\Gamma$ will mean a parameter $u$ such that the intermediate value $M(v, w)$ of (1.6) has the property

$$M(v, w) = (v + w)/2 \quad (2.1)$$

whenever $v, w \in [a, b]$; equivalently

$$0 = \mathbf{r}'(u) \times [\mathbf{r}(u + v) - \mathbf{r}(u - v)] \quad (2.2)$$

whenever $u \pm v \in [a, b]$. If $\Gamma$ has an additive parameter $u$, then $(u_0, \dots, u_n)$ is a locally optimal partition precisely when the $u_j$ form an arithmetic progression; that is, when

$$u_j - u_0 = (j/n)(b - a), \qquad j = 0, \dots, n. \quad (2.3)$$

However, $F_n$ achieves its global maximum; so this unique $(u_0, \dots, u_n)$ must yield the maximum.

Moreover, the existence of an additive parameter implies the validity of the subdivision algorithm: if $(u_0, \dots, u_n)$ is a locally optimal partition, and $(\bar{u}_0, \dots, \bar{u}_{2n})$ is the outcome when we apply this algorithm, then, by (2.1),

$$\bar{u}_j - \bar{u}_0 = (j/2n)(b - a), \qquad j = 0, \dots, 2n, \quad (2.4)$$

whence, by inspection $(\bar{u}_0, \dots, \bar{u}_{2n})$ is a locally optimal partition. In particular, the subdivision process yields inscribed polygons $K_n$ of maximal area whenever $n = 2^m$ and $m = 1, 2, \dots$. Therefore, we need only inquire which curves $\Gamma$ obey this extra assumption; the next two theorems answer this question.

For our introductory remarks, we have set the coordinate origin inside $K$; but clearly translations and invertible linear transformations of the $xy$-plane do not

change convexity and smoothness properties, or *ratios* of areas. Hence our arguments freely invoke the affine invariance of the problem to shift the origin, or to transform the coordinates, whenever this yields a simpler description of the curve $\Gamma$.

THEOREM 2.1. *If, in the plane, $\Gamma$ is an arc of any conic section, then $\Gamma$ has an additive parameter.*

*Proof.* Given any arc of (1) an ellipse, (2) a parabola, or (3) a hyperbola, apply some affine transformation to make it an arc of (1) the unit circle $x^2 + y^2 = 1$, (2) the standard parabola $y = x^2$, or (3) the first-quadrant hyperbola $xy = 1$. Respectively, these three curves have the additive parametrizations:

$$\begin{aligned} &(1)\ \mathbf{r}(u) = \mathbf{i}\cos u + \mathbf{j}\sin u, \\ &(2)\ \mathbf{r}(u) = \mathbf{i}u + \mathbf{j}u^2, \\ &(3)\ \mathbf{r}(u) = \mathbf{i}\exp u + \mathbf{j}\exp(-u). \end{aligned} \tag{2.5}$$

Indeed, the three alternatives (2.5) all satisfy a relation

$$\mathbf{r}(u+v) - \mathbf{r}(u-v) = g(v)\mathbf{r}'(u), \tag{2.6}$$

where, respectively, $g(v) = 2\sin v$, $2v$, or $2\sinh v$. However, (2.6) implies (2.2).

Earlier, we recalled that Archimedes, via the subdivision algorithm, inscribed a sequence of polygons in a parabolic arch. Directly, Theorem 2.1 shows that each of these polygons has maximal area for the given number of vertices or sides. Indeed, an analogous sequence of subdivision polygons has the same maximal property in an elliptical or hyperbolic arch. Moreover, the symmetry axes of the parabola, ellipse, or hyperbola may make an arbitrary angle with the base of the arch (e.g., Figure 1). Conversely, our next theorem, assuming less smoothness, asserts that *only* a conic section can have an additive parameter. Indeed, the functional equation (2.2) for such a parameter has the following partial derivative in the variable $v$:

$$0 = \mathbf{r}'(u) \times [\mathbf{r}'(u+v) + \mathbf{r}'(u-v)]. \tag{2.7}$$

Our argument will reduce (2.7) to the equation

$$\mathbf{r}'(u+v) + \mathbf{r}'(u-v) = g(v)\mathbf{r}'(u), \tag{2.8}$$

with an undetermined function $g(v)$; then each component of (2.8) is a previously studied equation of W. H. Wilson, and the solutions [Aczél, 1966, pp. 165–170] will yield the desired result.

THEOREM 2.2. *If $\Gamma$ has a $C^1$ additive parameter $u$ (where $u$ has domain $[a,b]$) such that $\mathbf{r}'(u) \neq 0$ and $\tau(u)$ is a strictly increasing continuous function, then $\Gamma$ is an arc of a conic section, and $\mathbf{r}(u)$ is essentially one of the parametrizations* (2.5).

*Proof.* Function $\mathbf{r}(u)$ has the property $\mathbf{r}'(u) \neq 0$; so (2.7) has the equivalent form

$$\mathbf{r}'(u+v) + \mathbf{r}'(u-v) = g(u,v)\mathbf{r}'(u), \tag{2.9}$$

where $g(u,v)$ is some continuous real-valued function. If overlapping subarcs of $\Gamma$ each have parametrizations of form (2.5), then $\Gamma$, by analytic continuation, has the *same* parametrization on all the subarcs. Hence we may suppose $\Gamma$ to be a short enough arc so that $\tau(b) - \tau(a) < \pi$. Then (2.9) has nonvanishing left side, whence $g(u,v) \neq 0$. If now we abbreviate

$$G(u,v) = g(u-v,v)g(u,v)g(u+v,v) - g(u-v,v) - g(u+v,v), \tag{2.10}$$

then direct substitution yields

$$\begin{aligned}
G(u,v)&[\mathbf{r}'(u-2v)+\mathbf{r}'(u+2v)]\\
&= G(u,v)g(u,2v)\mathbf{r}'(u)\\
&= g(u,2v)g(u-v,v)g(u+v,v)[\mathbf{r}'(u-v)+\mathbf{r}'(u+v)] \qquad (2.11)\\
&\quad - g(u,2v)[g(u-v,v)+g(u+v,v)]\mathbf{r}'(u)\\
&= g(u,2v)\{g(u+v,v)\mathbf{r}'(u-2v)+g(u-v,v)\mathbf{r}'(u+2v)\}.
\end{aligned}$$

If $v \neq 0$ then $\mathbf{r}'(u-2v)$ and $\mathbf{r}'(u+2v)$ are linearly independent vectors, and we may equate corresponding coefficients; thus

$$g(u,2v)g(u-v,v) = g(u,2v)g(u+v,v) = G(u,v). \qquad (2.12)$$

But $g(u-v,v) = g(u+v,v)$ by (2.12); so $g(u,v)$ has period $2v$ in its first argument. Also $g(u,2v) = G(u,v)/g(u+v,v)$ by (2.12), and $G(u,v)$ is a polynomial in functions $g(\cdot,v)$; so $g(u,2v)$ has period $2v$ in its first argument, or $g(u,v)$ has period $v$ in $u$. By induction, $g(u,v)$ has period $v \cdot 2^{-m}$ in $u$, where $m = 0,1,2,\ldots$ . By continuity, $g(u,v)$ is a constant function of $u$. Thus equation (2.9) takes the form (2.8). But the theory of W. H. Wilson's scalar equation [Aczél 1966, pp. 165–170] admits only the following nonzero solutions for this vector generalization:

$$\begin{aligned}
\mathbf{r}'(u) &= \mathbf{B}\cos\lambda u - \mathbf{A}\sin\lambda u,\\
\mathbf{r}'(u) &= \mathbf{A} + 2\mathbf{B}u, \qquad (2.13)\\
\mathbf{r}'(u) &= \mathbf{A}\exp\lambda u - \mathbf{B}\exp(-\lambda u),
\end{aligned}$$

where $\mathbf{A}$ and $\mathbf{B}$ are constant real vectors, while $\lambda$ is a nonzero real constant. Moreover, $\mathbf{A}$ and $\mathbf{B}$ must be linearly independent, because $\tau(u)$ is not constant. After an affine transformation we may take $\mathbf{A} = \mathbf{i}$, $\mathbf{B} = \mathbf{j}$; by a recalibration we may take $\lambda = 1$. Finally, integration yields the alternatives (2.5).

*Remark.* The authors have been unable to find whether, conversely, the validity of the subdivision algorithm implies the existence of an additive parameter. Indeed, if $u$ is an arbitrary $C^2$ parameter, then $M(v,v) = v$ and $M(v,w) = M(w,v)$. If also some function $\phi(u)$ were an additive parameter, then $M(v,w)$ would have the form

$$M(v,w) = \phi^{-1}[(\phi(v)+\phi(w))/2]. \qquad (2.14)$$

Either of the following additional relations implies the existence of such a function $\phi$ [Aczél 1966, pp. 278–299]:

$$\begin{aligned}
M(u, M(v,w)) &= M(M(u,v), M(u,w)),\\
M(M(t,u), M(v,w)) &= M(M(t,v), M(u,w)).
\end{aligned} \qquad (2.15)$$

However, the validity of the subdivision algorithm is the precise equivalent of a different relation:

$$M(v,w) = M[M(v, M(v,w)), M(M(v,w),w)]; \qquad (2.16)$$

and we have not been able to prove that (2.16) implies (2.14). (Added in proof: If $\Gamma$ is a $C^6$ curve then a somewhat tedious argument gives this result.)

**3. Rectangular Coordinates; Iterative Solution.** Initially, we introduced $x$ and $y$ axes, but so far we have not really used them. Here, under slight restrictions, we represent the curve $\Gamma$ in $xy$-coordinates, then give a test, and an algorithm, for locally optimal partitions, and discuss the convergence of this algorithm. Specifically, the affine invariance of our general problem allows a coordinatized formulation in any case where the parameter $u$ has domain $[a, b]$ and

$$\tau(b) - \tau(a) < \pi. \tag{3.1}$$

Indeed, then $\tau(b) - \tau(u) < \pi$ and $\tau(u) - \tau(a) < \pi$, so that

$$\begin{aligned} \mathbf{k} \times [\mathbf{r}'(a) - \mathbf{r}'(b)] \cdot \mathbf{r}'(u) = |\mathbf{r}'(u)|\{|\mathbf{r}'(a)|\sin(\tau(u) - \tau(a)) \\ + |\mathbf{r}'(b)|\sin(\tau(b) - \tau(u))\} > 0. \end{aligned} \tag{3.2}$$

If now we choose the unit vector $\mathbf{i}$ to parallel the (nonzero) vector $\mathbf{k} \times [\mathbf{r}'(a) - \mathbf{r}'(b)]$, then clearly $dx/du = \mathbf{i} \cdot d\mathbf{r}/du > 0$. Hence we may take the parameter $u$ to be the $x$-coordinate itself.

Accordingly, the curve $\Gamma$ will have parametrization $\mathbf{r}(x) = \mathbf{i}x + \mathbf{j}y(x)$, where $a \leqslant x \leqslant b$ and $y(x)$ is a $C^2$ function. Moreover, we may translate the origin so that $y(x)$ is a *nonnegative* function. If $(\cdot)' = (d/dx)(\cdot)$, then $y'(x) = \tan(\tau(x))$; thus $y'(x)$ is a strictly increasing $C^1$ function, and $-\pi/2 < \tau(x) < \pi/2$. Also, a partition $\Pi_n$ has form $(x_0, \dots, x_n)$, where $a = x_0 \leqslant x_1 \leqslant \cdots \leqslant x_{n-1} \leqslant x_n = b$; and the broken line $\Gamma_n$ has successive points $(x_0, y_0), \dots, (x_n, y_n)$, where

$$y_j = y(x_j), \qquad y_j' = y'(x_j), \qquad y_j'' = y''(x_j), \dots . \tag{3.3}$$

If $G/2$ (resp. $G_n/2$) is the area under $\Gamma$ (resp., $\Gamma_n$) but above the $x$-axis, then

$$G = 2\int_a^b y(x)\, dx, \quad G_n = x_n y_n - x_0 y_0 + \sum_{j=1}^{n} (x_j y_{j-1} - x_{j-1} y_j). \tag{3.4}$$

However, we wish to minimize the area $G_n - G$; therefore, we wish to minimize the area $G_n$.

By (3.4), this area is a function of $x_1, \dots, x_{n-1}$. Moreover,

$$\partial G_n/\partial x_j = (x_{j+1} - x_{j-1}) y_j' - (y_{j+1} - y_{j-1}), \qquad j = 1, \dots, n-1; \tag{3.5}$$

$$\partial^2 G_n/\partial x_j^2 = (x_{j+1} - x_{j-1}) y_j'', \qquad j = 1, \dots, n-1; \tag{3.6}$$

$$\partial^2 G_n/\partial x_j\, \partial x_{j-1} = \partial^2 G_n/\partial x_{j-1}\, \partial x_j = y_{j-1}' - y_j', \qquad j = 2, \dots, n-1; \tag{3.7}$$

and all other $\partial^2 G_n/\partial x_i\, \partial x_j = 0$, whence the Hessian $(\partial^2 G_n/\partial x_i\, \partial x_j)$ is a tridiagonal $(n-1) \times (n-1)$ matrix. Thus, if $D_0 = 1$ and otherwise $D_j$ is the determinant of the upper $j \times j$ submatrix, then expanding by minors yields the recursion

$$D_j = (x_{j+1} - x_{j-1}) y_j'' D_{j-1} - (y_{j-1}' - y_j')^2 D_{j-2}, \qquad j = 2, \dots, n-1. \tag{3.8}$$

Here (3.5) provides merely a restatement of our prior necessary conditions: if $(x_0, \dots, x_n)$ is a locally optimal partition, then the $x_j$ are distinct, and $\partial G_n/\partial x_j = 0$, $j = 1, \dots, n-1$. But (3.8) offers also a test, through well-known, sufficient conditions: namely $(x_0, \dots, x_n)$ *is* a locally optimal partition if the necessary conditions hold, *and* if $D_j > 0$, $j = 1, \dots, n-1$. More generally, the sequences $(x_0, \dots, x_n)$ subject only to the necessary conditions constitute the set $S$ of *stationary* partitions. No element of $S$ can yield a local *maximum* of $G_n$, because, from any given $(x_0, \dots, x_n)$ we can increase $G_n$ continuously until, for some $k$, we have $x_0 = \cdots$

$= x_{k-1}$ and $x_k = \cdots = x_n$. The next theorem further describes the set $S$. Clearly, we might obtain its result without $xy$-coordinates, hence, without the restriction (3.1).

THEOREM 3.1. *In $R^{n+1}$, the set $S$ of stationary partitions is the homeomorphic image of a closed subset $S_1$ in the compact interval $[a, b]$. If $y(x)$ is a $C^{m+1}$ function then $S$ is a $C^m$ image.*

*Proof.* At the endpoint $(b, y(b))$ we prolong $\Gamma$ slightly with, say, a short polynomial arc. Then we consider all sequences $(x_0, \ldots, x_n)$ such that $x_0 = a$ and $(x_n, y(x_n)) \in \Gamma$ but $x_n$ is not necessarily $b$, while

$$(x_{j+1} - x_{j-1}) y'(x_j) = y(x_{j+1}) - y(x_{j-1}), \qquad j = 1, \ldots, n-1. \tag{3.9}$$

If $x_1 = x_0$ then $x_n = x_0$; hence we suppose $x_1 > x_0$, and (3.9) implies $x_n > x_{n-1} > \cdots > x_0$. Thus each $x_{j+1}$, by (3.9), is a $C^m$ function of $x_j$ and $x_{j-1}$. But we have fixed $x_0$; so, inductively, each $x_j$ is a $C^m$ function of $x_1$, whereas, clearly, $x_1$ is a $C^m$ function of $(x_0, \ldots, x_n)$. Furthermore, $(x_0, \ldots, x_n)$ is a stationary partition precisely when $x_1 \in S_1$, where $S_1 = \{x_1 \in [a, b] : x_n(x_1) = b\}$.

Furthermore, a simple iterative algorithm yields successive partitions $(x_0, \ldots, x_n)$ that more and more nearly satisfy the necessary conditions. Indeed, if $\Pi_n$ is any nondegenerate partition, and $Mx_j$ abbreviates $M(x_{j-1}, x_{j+1})$, so that

$$(x_{j+1} - x_{j-1}) y'(Mx_j) = y_{j+1} - y_{j-1}, \tag{3.10}$$

then $M_j\Pi_n$ will denote the modified partition where $Mx_j$ replaces $x_j$ but all other $x_i$ remain the same. In other words, $M_j\Pi_n$ optimizes $x_j$ relative to its nearest neighbors. Then Taylor's theorem with remainder implies that

$$G_n(\Pi_n) - G_n(M_j\Pi_n) = \frac{1}{2}(x_{j+1} - x_{j-1})(x_j - Mx_j)^2 y''(\lambda Mx_j + (1-\lambda)x_j) \tag{3.11}$$

for some $\lambda$, where $0 < \lambda < 1$. Either the geometry or (3.11) shows that $G_n(M_j\Pi_n) \leqslant G_n(\Pi_n)$. Moreover, if $y''(x) \geqslant \gamma > 0$ when $a \leqslant x \leqslant b$, and if $\varepsilon_j = (2/\gamma)[G_n(\Pi_n) - G_n(M_j\Pi_n)]$, then

$$|Mx_j - x_j| \leqslant \varepsilon_j^{1/3}, \tag{3.12}$$

since either $|x_{j+1} - x_{j-1}| \leqslant \varepsilon_j^{1/3}$ or, otherwise, (3.11) implies that

$$G_n(\Pi_n) - G_n(M_j\Pi_n) > (\gamma/2)\varepsilon_j^{1/3}(Mx_j - x_j)^2. \tag{3.13}$$

Our proposed algorithm merely repeats this process. In any fixed order, it optimizes *every* $x_j$ relative to its nearest neighbors. The initial $\Pi_n^{(0)}$ may be any nondegenerate partition. If $\Pi_n^{(k)}$ is any later iterate, and $p$ is any permutation of $(1, \ldots, n-1)$, then we replace successively each $x_{p(j)}$, where $j = 1, \ldots, n-1$, so that the next iterate in our sequence is the partition

$$\Pi_n^{(k+1)} = M_{p(n-1)} \cdots M_{p(2)} M_{p(1)} \Pi_n^{(k)}. \tag{3.14}$$

Section 1 observes that, within some visual accuracy, we can perform each step $M_j$ with a straightedge; hence this algorithm has direct geometric meaning. Moreover, we can treat rather fully its convergence.

THEOREM 3.2. *If $y''(x) \geqslant \gamma > 0$ when $a \leqslant x \leqslant b$, then the accumulation points of the proposed sequence $\{\Pi_n^{(k)}\colon k = 0, 1, 2, \dots\}$ form a nonvoid compact connected subset of the stationary partitions. On this subset the function $G_n$ is constant.*

*Proof.* The function $G_n$ has a compact domain; hence the accumulation points form a nonvoid compact subset $S^*$. However, $\{G_n(\Pi_n^{(k)})\colon k = 0, 1, 2, \dots\}$ is a nonincreasing sequence with lower bound $G$; so it has a limit $G^*$, which, by continuity, must be the value $G_n(\Pi_n)$ everywhere on $S^*$. If $\Pi_n^{(k)} = (x_0^{(k)}, \dots, x_n^{(k)})$, and $\Pi_n^{(k+1)} = (x_0^{(k+1)}, \dots, x_n^{(k+1)})$, then

$$\begin{aligned} \|\Pi_n^{(k+1)} - \Pi_n^{(k)}\|_3 &= \left\{ \sum_{j=0}^{n} |x_j^{(k+1)} - x_j^{(k)}|^3 \right\}^{1/3} \\ &\leqslant \left\{ \sum_{j=1}^{n-1} \varepsilon_{p(j)} \right\}^{1/3} = (2/\gamma)^{1/3} \left\{ G_n\big(\Pi_n^{(k)}\big) - G_n\big(\Pi_n^{(k+1)}\big) \right\}^{1/3}. \end{aligned} \tag{3.15}$$

Any connected component of $S^*$ is a compact subset of $R^{n+1}$. Thus $S^*$ cannot have disjoint components; indeed, two such components would have positive minimum distance, whereas the right side of (3.15) approaches zero, so that the distance between successive $\Pi_n^{(k)}$ approaches zero. Also by (3.15), the quantities $|x_j^{(k)} - Mx_j^{(k)}|$ approach zero, whence, by continuity, the elements of $S^*$ are stationary partitions.

Example 4.1 specifies a curve $\Gamma$ such that $G_3$ has two local minima and a saddle point. Thus, conceivably, an accumulation point of the $\Pi_n^{(k)}$ might be a saddle point of $G_n$; but the corresponding iteration would be an unstable process, because each nontrivial step must decrease $G_n$. Hence perturbations by roundoff error would probably redirect the $\Pi_n^{(k)}$ toward a local minimum. If a function has no unique local minimum, then some authors [e.g., Wolfe 1969] describe an optimization algorithm as convergent when the gradients approach zero on the resulting sequence. Theorem 3.2 yields more; it implies that the gradients $(\partial G_n/\partial x_1, \dots, \partial G_n/\partial x_{n-1})(\Pi_n^{(k)})$ approach zero as $k \to \infty$, and that the $\Pi_n^{(k)}$ themselves have a limit whenever the set $S$ of stationary partitions includes no connected set larger than a single point. By Theorem 3.1, a component of $S$ is either a point or a $C^1$ arc. Thus the $\Pi_n^{(k)}$ have a limit when $S$ is any finite or denumerable set.

Example 4.2 presents a curve $\Gamma$ such that $S$ includes a $C^1$ arc (and, indeed, the $\Pi_n^{(k)}$ of iteration (3.14) would *not* have a limit if they pursued a wandering point in such an arc). But even near such arcs, we can obtain a stronger convergence result from smooth enough curves: if $y(x)$ has *four* continuous derivatives, then theorems of discrete dynamics [Aulbach 1984; Hirsch, Pugh, and Shub 1977] and computational linear algebra [Young 1971] guarantee a limit for the defined sequence $\{\Pi_n^{(k)}\}$. The proof [Lew and Quarles 1988] involves enough details so that we omit it here, but the result stretches to cover Example 4.2, whence, there also, the iteration has a limit.

**4. Examples with Multiple Extrema.** A curve $\Gamma$ may satisfy all the assumptions of Theorem 3.2, and still it may have an unexpectedly large set of locally optimal partitions. Here, in $xy$-coordinates, we give two examples that show the possibilities.

In each case, the polygonal line approximating the curve has only three segments, i.e., two adjustable vertices. In the first example, $G_3$ has two local minima, whence $F_3$, by our previous remarks, has two local maxima. In the second, the minima form a $C^\infty$ arc; so the curve $\Gamma$ has a *continuum* of locally optimal partitions.

No doubt, asymmetric curves would yield the same pathologies, but, in these examples, for simplicity, our $y(x)$ will be an *even* function. Then $y(x) = f(|x|)$ for some function $f$, and in each case we shall first specify this $f$. In both examples, $y(x)$ will be a piecewise-rational function that has two continuous derivatives even at the knots, i.e., the transition points. (Indeed, a bit more work would produce even smoother transitions.) In Example 4.1, each piece will be a cubic polynomial, whence, by definition, $y(x)$ is a $C^2$ cubic spline [e.g., de Boor 1978]. In Example 4.2, four pieces will be cubic polynomials, while two others are simple rational functions.

**Example 4.1 (Figure 2).** In general, a curve $\Gamma$ may have peculiar properties if, despite two continuous derivatives, it lies near—and just rounds slightly—some polygonal line. The dashed line in Figure 2 is the polygonal line behind this example; it has vertices $(0,0)$, $(\pm 6,3)$, $(\pm 8,5)$, $(\pm 9,12)$. Indeed, one checks easily that, on $[0, +\infty)$, the following values define a $C^2$ cubic spline $f(x)$ with knots 0, 2, 6, 7.

| $x$ | $f(x+)$ | $f'(x+)$ | $f''(x+)$ | $f'''(x+)$ |
|---|---|---|---|---|
| 0 | 0.35 | 0.0 | .475 | $-.225$ |
| 2 | 1.0 | 0.5 | .025 | .00375 |
| 6 | 3.24 | 0.63 | .04 | .66 |
| 7 | 4.0 | 1.0 | .7 | 3.45 |

(4.1)

The corresponding curve intersects the dashed line when $x = 2, 7, 9$.

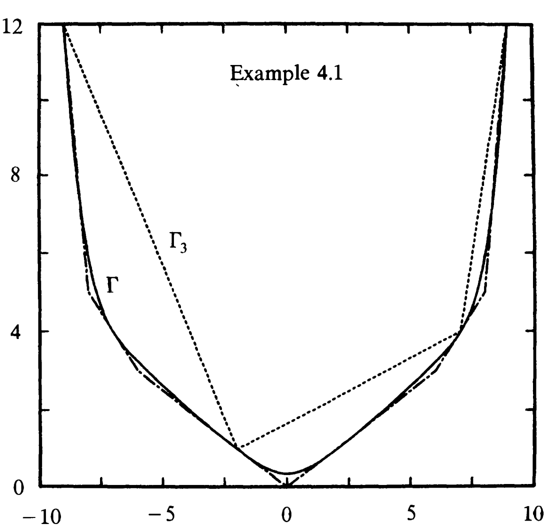



FIG. 2. The solid curve $\Gamma$ has two locally optimal inscribed polygons $\Gamma_3$; we show one.

Thus if one defines $y(x) = f(|x|)$, then, on $(-\infty, +\infty)$, the resulting $y(x)$ is a $C^2$ cubic spline with knots, $0, \pm 2, \pm 6, \pm 7$. Moreover, $y(\pm 9) = 12$, and $y''(x)$ is a strictly positive function. Now let $[a, b] = [-9, 9]$, and consider the partition $(x_0, x_1, x_2, x_3) = (-9, -2, 7, 9)$. The table (4.1) shows that $Mx_1 = x_1$ and $Mx_2 = x_2$, where (3.10) defines $Mx_j$. Also

$$\left(\partial^2 G_3/\partial x_i\, \partial x_j\right) = \begin{pmatrix} 0.4 & -1.5 \\ -1.5 & 7.7 \end{pmatrix}, \tag{4.2}$$

and this is a positive definite matrix. Therefore the partition $(-9, -2, 7, 9)$ is a local minimum of $G_3$; so likewise its reflection $(-9, -7, 2, 9)$ is a local minimum of $G_3$. Indeed, a *third* partition satisfies the necessary conditions, namely, $(-9, -x_2, x_2, 9)$ where $x_2 \simeq 5.62235$; but further analysis shows that this is a saddle point.

**Example 4.2 (Figures 3 and 4).** Here we seek a function $y(x)$ on an interval $[a, b]$ and a continuous family $A$ of sequences $(x_0, x_1, x_2, x_3)$ such that $(x_0, x_3) = (a, b)$ and every sequence is a stationary partition, i.e.,

$$(x_2 - x_0)y_1' = y_2 - y_0, \qquad (x_3 - x_1)y_2' = y_3 - y_1. \tag{4.3}$$

We may suppose that $x_1 = h(x_2)$, where $h$ is an increasing continuous function. If $y_1(x_1)$, $y_2(x_2)$ are independent functions, and we put

$$\begin{aligned} \xi_1 &= (x_3 - x_1)^{-1}, & \eta_1 &= (y_1 - y_3)/(x_3 - x_1), \\ \xi_2 &= (x_2 - x_0)^{-1}, & \eta_2 &= (y_2 - y_0)/(x_2 - x_0), \end{aligned} \tag{4.4}$$

then $h$ specifies a decreasing continuous function $\phi$: $\xi_2 \to \xi_1$, and (4.3) becomes $\eta_1 + \eta_2 = \xi_1\, d\eta_1/d\xi_1 = \xi_2\, d\eta_2/d\xi_2$. Eliminating $\eta_1$ yields a second-order differential equation that has general solution

$$\eta_1 = C\xi_1\xi_2 - \eta_2, \qquad \eta_2 = C\left[\gamma_0 + \int^{\xi_2}\phi(\xi)\, d\xi\right], \tag{4.5}$$

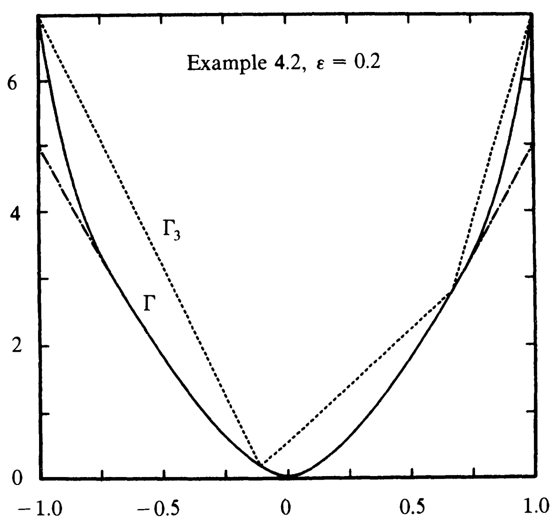



FIG. 3. The solid curve $\Gamma$ has a continuum of locally optimal inscribed polygons $\Gamma_3$; we show one.

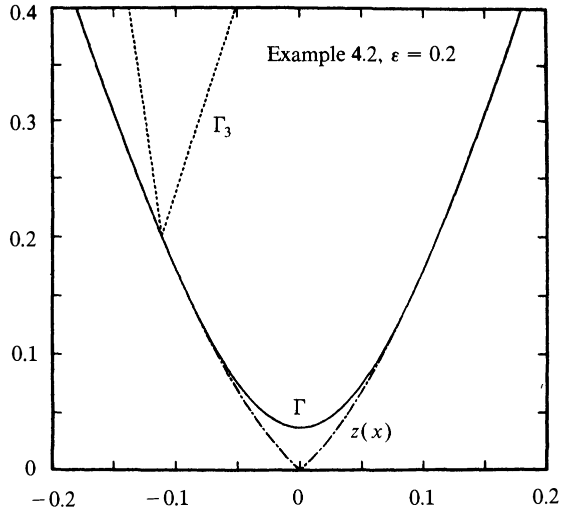



FIG. 4. Detail of Figure 3. Solid and dashed curves plot, resp., $y(x)$ and $z(x)$.

where $\gamma_0$ and $C$ are integration constants. Many functions $\phi$ generate solutions (4.5), but the resulting $y_1(x_1)$, $y_2(x_2)$ do not obey all the boundary conditions. Hence our example modifies one such solution in certain neighborhoods of the boundary points.

Indeed, here we take $[a, b] = [-1, 1]$, and we choose $x_1 = h(x_2)$ so that $(1 - 3x_1)(1 + 3x_2) = 4$. Then (4.5) yields a particular solution $y_2(x_2) = g(x_2)$, where $g(x) = x(1 + 9x)/(1 + x)$. This $g$, modified near 0 and 1, gives the function $f$ used in our example. Specifically, we let $0 < \varepsilon < 1/2$ and $c = \varepsilon/(2 - \varepsilon)$ and $d = (1 - \varepsilon)/(1 + \varepsilon)$. Then $0 < c < d < 1$; and moreover $-d \leqslant h(x_2) \leqslant -c$ precisely when $c \leqslant x_2 \leqslant d$. On $[c, d]$, for any such $\varepsilon$, we let $f(x) = g(x)$. On $[0, c)$ we let $f(x)$ be the cubic polynomial such that $(f(c), f'(c), f''(c)) = (g(c), g'(c), g''(c))$ and $f'(0) = 0$. On $(d, +\infty)$ we let $f(x)$ be the cubic polynomial such that $(f(d), f'(d), f''(d)) = (g(d), g'(d), g''(d))$ and $f(1) = 7$. These values imply that $f(0) > 0$ and $f''(x) \geqslant f''(d) > 0$. Finally, we set $y(x) = f(|x|)$, so that $y(\pm 1) = 7$, and, on $(-\infty, +\infty)$, the defined $y(x)$ is a $C^2$ convex positive function with strictly positive $y''(x)$. Also, for comparison, we set $z(x) = g(|x|)$. In Figures 3 and 4, the solid and dashed lines plot, respectively, $y(x)$ and $z(x)$ for $\varepsilon = 0.2$.

Clearly, the following set is a $C^\infty$ arc:

$$A = \{(x_0, x_1, x_2, x_3) \colon -d \leqslant x_1 \leqslant -c;\ c \leqslant x_2 \leqslant d;\ (1 - 3x_1)(1 + 3x_2) = 4\}. \tag{4.6}$$

Moreover, by the preceding argument, or by direct calculation, we find that any $(x_0, x_1, x_2, x_3)$ in $A$ satisfies the requirements (4.3) for a stationary partition. Though the function $y(x)$ has just two continuous derivatives, still the iteration (3.14) approaches a limit. Indeed, if finitely many steps either take $x_1$ into $[-d, -c]$ or take $x_2$ into $[c, d]$, then just one more step achieves the limit. Otherwise $x_1$ and $x_2$, after finitely many steps, never leave intervals in which $y(x)$ is a cubic polynomial. But the last paragraph of Section 3 cites a result for $C^4$ functions $y(x)$ that now shows the existence of the limit.

Finally, the elements of $A$ are all (nonstrict) local minima of $G_3$. Indeed, $G_3$, by Theorem 3.2, has some constant value $G^*$ on $A$ and $G^* \leqslant G_3(x_0, x_1, x_2, x_3)$ whenever a change in just one $x_j$ can take $(x_0, x_1, x_2, x_3)$ into $A$. We must prove this inequality also when $(x_0, x_1, x_2, x_3)$ is a partition sufficiently near an endpoint of $A$, but when $(x_0, x_1, x_2, x_3)$ does not obey the preceding condition. By symmetry, we need consider only regions, for small positive $\delta$, where $-c < x_1 < -c + \delta$ and $d < x_2 < d + \delta$. But if $G_3$ becomes $H_3$ when $y(x)$ becomes $z(x)$, then, in any such region, $G^* \leqslant H_3(x_0, x_1, x_2, x_3)$, so we need show only that $H_3 \leqslant G_3$. However, the Taylor series for $G_3 - H_3$ at $(-c, d)$ has positive third-degree terms in the stated regions, and has no terms of lower degree. Hence $H_3 \leqslant G_3$ for small enough $\delta$.

**5. Criteria for Unique Solution.** We have just seen two examples with multiple solutions, but we can now give a condition for uniqueness. Again let $u$ be any admissible curve parameter, and let $v_0, v_1, \ldots, v_m$ be any increasing $u$-values such that $v_j = M(v_{j-1}, v_{j+1})$ when $j = 1, \ldots, m-1$. Here we allow $v_0 = a = -\infty$, and we define $\Delta v_j = v_j - v_1$, where $j = 0, \ldots, m$. If $v_0 = -\infty$ and $v_1$ has any finite value, then $\Delta v_0$ will have the "fixed" value $-\infty$. Geometric arguments show that any two of the values $v_{j-1}, v_j, v_{j+1}$, uniquely determine the third; by induction, this implies that $v_0$ and $v_1$ uniquely determine $v_m$. Thus we may ask whether $u$ has the following properties.

$P_0(m)$: If $\Delta v_0$ is fixed, then $\Delta v_m$ is a constant function of $v_1$.
$P_1(m)$: If $\Delta v_0$ is fixed, then $\Delta v_m$ is a nondecreasing function of $v_1$.
$P_2(m)$: If $v_1$ is fixed, then $v_m$ is a strictly decreasing function of $v_0$.
$P_3(m)$: If $v_0$ is fixed, then $v_m$ is a strictly increasing function of $v_1$.

Geometric arguments show easily that *all* parameters $u$ have properties $P_2(2)$ and $P_3(2)$. Hence we call $u$ an *invariant* or a *monotone* parameter if, respectively, it has property $P_0(2)$ or $P_1(2)$. Clearly, any invariant parameter is a monotone parameter.

THEOREM 5.1. *If $u$ is a monotone parameter, then $u$ has properties $P_1(m)$, $P_2(m)$, $P_3(m)$, where $m = 2, 3, 4 \ldots$.*

*Proof* (sketch). We have $P_1(2)$, $P_2(2)$ and $P_3(2)$. Hence we assume $P_1(m-1)$, $P_2(m-1)$, $P_3(m-1)$ for an induction step, and we attack $P_2(m)$, $P_1(m)$, $P_3(m)$ —in that order. To prove each $P_j(m)$, we must compare sequences $(v_0^{(1)}, \ldots, v_m^{(1)})$ and $(v_0^{(2)}, \ldots, v_m^{(2)})$ with certain assumptions on their first two components. To make each comparison, we define an intermediate $m$-tuple or $(m+1)$-tuple with some properties of each given sequence, and we deduce chained inequalities.

COROLLARY 5.2. *If $\Gamma$ has an invariant (or monotone) parameter $u$, then our problem has a unique solution.*

*Proof.* If $u$ has domain $[a, b]$, and $v_0 = a$, then $v_m$, by $P_3(m)$, is a strictly increasing function of $v_1$. Hence, only a single $v_1$ can make $v_m = b$.

If we have a continuous family $\{T(u)\colon a \leqslant u \leqslant b\}$ of plane affine transformations such that $T(u_1 + u_2) = T(u_1)T(u_2)$ for all admissible $u_1, u_2, u_1 + u_2$, then we can find a curve $\Gamma$ with an invariant parameter. We simply let $\mathbf{r}(u) = T(u)\mathbf{q}$, where $a \leqslant u \leqslant b$, for any vector $\mathbf{q}$ that is not a fixed point of the operators $T(u)$. Then $\mathbf{r}(u)$ traces a curve $\Gamma$, and $u$ is its invariant parameter. This remark underlies the next example, but we give only a direct verification.

**Example 5.3.** If $(x, y)$ and $(r, \theta)$ are, respectively, rectangular and polar coordinates, then the following curves have invariant parameters:

1. any power-law curve $y = x^{\alpha}$, $0 < x < +\infty$;
2. the special curve $y = x \cdot \log x$, $0 \leqslant x < +\infty$;
3. the exponential curve $y = \exp x$, $-\infty < x < +\infty$;
4. the logarithmic spiral $r = \exp \alpha\theta$, $-\infty < \theta < +\infty$.

Here $\alpha$ is any real constant, and the power-law domain includes 0 whenever $\alpha > 0$. Indeed, we introduce directly the parameters: (1 or 2) $u = \log x$; (3) $u = x$; (4) $u = \theta$. Then we need only verify the relation $h(v_0 - v_1) = h(v_2 - v_1)$, where, respectively: (1) $h(u) = \exp \alpha u - \alpha \cdot \exp u$; (2) $h(u) = (1 - u) \cdot \exp u$; (3) $h(u) = -u + \exp u$; (4) $h(u) = (\alpha \sin u - \cos u)\exp u$. Thus our problem has a unique solution if any listed curve includes the specified $\Gamma$. The power-laws and logarithmic spirals, for suitable values $\alpha$, yield the conic sections in the standard form (2.5). Moreover, the image of a listed curve under any affine transformation is an admissible curve; for example, the power-laws yield the curves with equations $y = x + x^{\alpha}$.

Further, via Corollary 5.2, we can admit some curves *without* invariant parameters. Here, for convenience, we again use $x$ as the parameter, and let $\mathbf{r}(x) = \mathbf{i}x + \mathbf{j}y(x)$ for some $C^2$ function $y(x)$. Once more we add the hypothesis that $y''(x) \geqslant \gamma > 0$ when $a \leqslant x \leqslant b$. If $a \leqslant x_0 < x_1 < x_2 \leqslant b$ and $x_1 = M(x_0, x_2)$ then $y(x_2) - y(x_0) = (x_2 - x_0)y'(x_1)$, which takes the compact form

$$g(x_0, x_1) = g(x_2, x_1) \tag{5.1}$$

when $g(x, x_1) = y(x) - x \cdot y'(x_1)$. Differentiating this yields the relations

$$\partial(x_2 - x_1)/\partial(x_1)|_{x_1 - x_0 = \text{const.}} = [g^*(x_0, x_1) - g^*(x_2, x_1)]/[y'(x_2) - y'(x_1)], \tag{5.2}$$

$$\partial(x_1 - x_0)/\partial(-x_1)|_{x_2 - x_1 = \text{const.}} = [g^*(x_0, x_1) - g^*(x_2, x_1)]/[y'(x_1) - y'(x_0)], \tag{5.3}$$

where $g^*(x, x_1) = y'(x) - x \cdot y''(x_1)$. Clearly, (5.2) becomes zero precisely when

$$g^*(x_0, x_1) = g^*(x_2, x_1). \tag{5.4}$$

However, $\Gamma$ satisfies the uniqueness condition $P_1(2)$ if (5.2) is always positive. (Then (5.3) is also positive; a reversed parameter $-x$ yields exactly the same examples.)

If $y(x)$ is an analytic function and $x_1 - x_0$ is a small enough quantity, then, by the implicit function theorem, $x_2 - x_1$ is an analytic function of $x_1 - x_0$. Indeed, one finds a Taylor expansion

$$x_2 - x_1 = (x_1 - x_0) - \rho(x_1) \cdot (x_1 - x_0)^2 + \rho(x_1)^2(x_1 - x_0)^3 + O\big((x_1 - x_0)^3\big), \tag{5.5}$$

where $\rho(x) = y'''(x)/3y''(x)$. Thus

$$\partial(x_2 - x_1)/\partial x_1|_{x_1 - x_0 = \text{const.}} = -(d\rho/dx_1) \cdot (x_1 - x_0)^2 + O\big((x_1 - x_0)^3\big) \tag{5.6}$$

when $x_1 - x_0 \to 0$. This fact about small $x_1 - x_0$ implies a result for arbitrary $x_1 - x_0$.

THEOREM 5.4. *Let $y(x)$ be a real-analytic function and $y''(x) \geqslant \gamma > 0$ when $a \leqslant x \leqslant b$. Let $d\rho/dx < 0$ when $a < x < b$, where $\rho(x) = y'''(x)/3y''(x)$. If* (5.1) *and* (5.4) *have no common solution, then $x$ is a monotone parameter; so our problem has a unique solution.*

*Proof.* If $d\rho/dx < 0$ then small enough $x_1 - x_0$ yield positive (5.2). If (5.1) and (5.4) have no common solution then (5.2) cannot be zero.

**Example 5.5.** If $y(x) = -\cos x$, where $-\pi/2 \leqslant x \leqslant \pi/2$, then $3\rho(x) = -\tan x$, which is a decreasing function. If (5.1) and (5.4) have a common solution $(x_0, x_1, x_2)$, then

$$-g(x_0, x_1) + ig^*(x_0, x_1) = -g(x_2, x_1) + ig^*(x_2, x_1), \tag{5.7}$$

where $g(x, x_1) = -\cos x - x \cdot \sin x_1$ and $g^*(x, x_1) = \sin x - x \cdot \cos x_1$. Thus

$$\exp[i(x_0 - x_1)] - i(x_0 - x_1) = \exp[i(x_2 - x_1)] - i(x_2 - x_1). \tag{5.8}$$

Now, the real and imaginary parts show that (5.8) has no solution where $-\pi/2 \leqslant x_0 < x_1 < x_2 \leqslant \pi/2$.

**6. Approximate Solution; Numerical Examples.** Only a conic section can have an additive parameter $u$, i.e., one that satisfies (2.2). However, any admissible $\Gamma$ can have an "infinitesimally additive" parameter that more and more *nearly* satisfies (2.2) as $v \to 0$. Indeed, first let $\mathbf{r}(x) = \mathbf{i}x + \mathbf{j}y(x)$, where $y(x)$ is a $C^4$ function. Then Taylor's theorem implies that

$$\begin{aligned}
&[y(x(u+v)) - y(x(u-v))]/2 \\
&\quad = v \cdot y'(x(u)) \cdot x'(u) + (v^3/6)\big[y'''(x(u))x'(u)^3 \\
&\qquad + 3y''(x(u))x'(u)x''(u) + y'(x(u))x'''(u)\big] + o(v^3),
\end{aligned} \tag{6.1}$$

$$\begin{aligned}
y'(x(u))[x(u+v) - x(u-v)]/2 &= v \cdot y'(x(u)) \cdot x'(u) \\
&\quad + (v^3/6)[y'(x(u))x'''(u)] + o(v^3),
\end{aligned} \tag{6.2}$$

and these expansions have the same first term. An additive parameter $u$ would imply precisely equal expansions. Thus we require equal $v^3$-terms, and this yields the relation

$$0 = y'''(x(u)) \cdot x'(u)^3 + 3y''(x(u)) \cdot x'(u)x''(u). \tag{6.3}$$

But $x'(u) \neq 0$ and $y''(x) \neq 0$; so (6.3) has the first integral

$$du/dx = C \cdot |y''(x)|^{1/3}, \qquad C = \text{constant}. \tag{6.4}$$

If $s$ = arc length and $\kappa$ = curvature, then (6.4) has the Euclidean-invariant form

$$du/ds = C \cdot |\kappa(s)|^{1/3}, \qquad C = \text{constant}. \tag{6.5}$$

However, we can use (6.5) even when curve $\Gamma$ has no *global* representation $\mathbf{r}(x) = \mathbf{i}x + \mathbf{j}y(x)$, because this discussion needs only *local* properties. Indeed, we can use (6.5) to define a parameter $u$ even when $\kappa(s)$ is merely a continuous function, or has isolated zeros. Furthermore, equally spaced $u$-values locate the optimal partition when $u$ is an *additive* parameter; so *here* equally spaced $u$-values should produce good approximate solutions $\Pi_n$ as $n \to \infty$.

However, Section 3 has described a convergent iteration that yields locally optimal partitions, and any such process will obtain a faster result when it has a good starting point. Also, Section 5 has shown that certain curves $\Gamma$ have unique optimal partitions. Accordingly, for various such curves, we have obtained an initial guess $(x_0^{(0)}, \ldots, x_n^{(0)})$ via (6.4), computed the optimal partition $(x_0^{(\infty)}, \ldots, x_n^{(\infty)})$ by our algorithm, and studied the accuracy of the approximate starting point. In each case, the curve $\Gamma$ has equation $y = f(x)$ on an interval $[a, b]$.

We report first our numerical observations for some particularly "good" curves, namely, those where $f(x)$ satisfies all hypotheses of Theorem 3.2, and where the iterates, as a result, have a guaranteed limit. Such curves $\Gamma$ have strictly positive curvature at each point, because $f''(x) \geqslant \gamma > 0$ on $[a, b]$. Indeed, if $f(x) = \exp x$ and $[a, b] = [0, 1]$, or if $f(x) = x^\alpha$ and $[a, b] = [1, 2]$, then our computations suggest

$$\max_j |x_j^{(\infty)} - x_j^{(0)}| \sim \text{constant} \cdot n^{-2} \quad \text{as } n \to +\infty. \tag{6.6}$$

Hence, for such "good" curves generally, (6.6) may characterize the accuracy of the approximation.

On the other hand, if $f(x) = x^\alpha$ and $[a, b] = [0, 1]$, then $f''(a) = 0$ when $\alpha > 2$, and thus $\Gamma$ has *zero* curvature at its left endpoint. For such "less good" curves, Theorem 3.2 does not insure convergence, but, empirically, the iterates still approach a limit, and the data now suggest

$$\max_j |x_j^{(\infty)} - x_j^{(0)}| \sim \text{constant} \cdot n^{-3/(1+\alpha)} \quad \text{as } n \to +\infty. \tag{6.7}$$

Similarly, if $f(x) = -\cos x$ and $[a, b] = [0, \pi/2]$, then $f''(b) = 0$ but $f'''(b) \neq 0$. Moreover, an equivalent curve under affine transformation has $f(x) = x - \sin x$ on $[0, \pi/2]$. The data for either curve fit a relation consistent with (6.7):

$$\max_j |x_j^{(\infty)} - x_j^{(0)}| \sim \text{constant} \cdot n^{-3/4} \quad \text{as } n \to +\infty. \tag{6.8}$$

For these "less good" curves, one may also ask how the initial guess anticipates the worst final result. Small values of the integer $n$ show the typical behavior of the approximation. Indeed, if $f(x) = x^{12}$ and $[a, b] = [0, 1]$, while $n = 2$ or 3, then $x_1^{(0)} - x_1^{(\infty)} \simeq .06$, since

$$\begin{aligned} n = 2&: x_1^{(0)} \simeq .852180, \ x_1^{(\infty)} \simeq .797797 \\ n = 3&: x_1^{(0)} \simeq .776060, \ x_1^{(\infty)} \simeq .711537 \\ &\quad x_2^{(0)} \simeq .910675, \ x_2^{(\infty)} \simeq .891877. \end{aligned} \tag{6.9}$$

More generally, the greatest cause of inaccuracy of the initial guess is apparently to overestimate the distance between points of zero curvature and the nearest distinct positions $x_j$. Still, if we let $f(x) = x^\alpha$ on $[0, 1]$, and if, for each $n$, we choose the exponent $\alpha$ yielding the poorest outcome, then the largest such errors seem to have a bound .07. Moreover, these worst exponents seem all to exceed 12. Hence errors of this size require a point of zero curvature where the tangent to the curve has rather high order of contact. We obtain better guesses for smaller exponents $\alpha$, i.e., lower orders of contact: if $f(x) = x^4$ and $[a, b] = [0, 1]$, then

$$\begin{aligned} n = 2&: x_1^{(0)} \simeq .659754, \ x_1^{(\infty)} \simeq .629961 \\ n = 3&: x_1^{(0)} \simeq .517282, \ x_1^{(\infty)} \simeq .486038 \\ &\quad x_2^{(0)} \simeq .784053, \ x_2^{(\infty)} \simeq .771537. \end{aligned} \tag{6.10}$$

If $f(x) = x \cdot \log x$ and $[a, b] = [0, 1]$, then the origin is again a potentially troublesome point, because the curve there has *infinite* curvature. However our algorithm still yields a numerical result, and, in this case, the initial guess slightly *underestimates* the distance between the origin and $x_1^{(\infty)}$. The algorithm of Section 3 is stable, reliable, and geometrically meaningful, but its convergence in these examples, while linear, is slow. No doubt a more sophisticated scheme would produce faster answers; however, *guaranteed* convergence might then require additional safeguards.

**Acknowledgments.** The authors wish to thank Michael Shub and Philip Wolfe for providing references and examples that clarify the convergence properties of iterative algorithms.

REFERENCES

1. J. Aczél, Lectures on Functional Equations and Their Applications, Academic Press, New York, 1966.
2. B. Aulbach, Continuous and Discrete Dynamics near Manifolds of Equilibria, Lecture Notes in Mathematics 1058, Springer-Verlag, New York, 1984, Theorem 7.1.
3. C. De Boor, A Practical Guide to Splines, Applied Mathematical Sciences 27, Springer-Verlag, New York, 1978.
4. C. H. Dowker, On minimum circumscribed polygons, *Bull. AMS*, 50 (1944), 120–122.
5. T. L. Heath, The Works of Archimedes, Cambridge Univ. Press, London, 1912; Dover Publications, New York.
6. M. W. Hirsch, C. C. Pugh, and M. Shub, Invariant Manifolds, Lecture Notes in Mathematics 583, Springer-Verlag, New York, 1977, Theorem 4.1.
7. J. S. Lew and D. A. Quarles, Jr., Convergence of an iterative algorithm for optimal inscribed polygons, IBM Research Report RC 13882, 20 July 1988.
8. D. E. McClure and R. A. Vitale, Polygonal approximation of plane convex bodies, *J. Math. Anal. and Appl.*, 51 (1975) 326–358.
9. L. F. Tóth, Approximation by polygons and polyhedra, *Bull. AMS*, 54 (1948) 431–438.
10. P. Wolfe, Convergence conditions for ascent methods, *SIAM Review*, 11 (1969) 226–235.
11. D. M. Young, Iterative Solution of Large Linear Systems, Academic Press, New York, 1971, Sections 5.2–5.4.

# UNSOLVED PROBLEMS

EDITED BY RICHARD GUY

*In this department the* MONTHLY *presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial results. Manuscripts should be sent to Richard Guy, Department of Mathematics and Statistics, The University of Calgary, Calgary, Alberta, Canada T2N 1N4.*

## Unsolved Problems Come of Age

RICHARD K. GUY

This article completes 21 years of what Victor Klee started as the Research Problems section. In December, 1970 he asked if I would take it over. Two or three generations of editors later we realized that it should be called the Unsolved Problems section. In December, 1971 we jointly published the first of the two yearly updating articles, of which this is the tenth. Regular readers will know that references in brackets are year and page numbers of this MONTHLY, while dates in parentheses refer to publications listed at the end, and other items are labelled (tbp) or (wrc) according as they are likely to be published formally, or remain as written communications.

I'm not aware of any progress with the Kotzig-Ringel tree-labelling problem [1969, 1128] but Joseph Gallian (tbp) has a survey paper on graph labelling problems which contains a bibliography of 78 items, 20 of which were not on my earlier list of 147. Also, expect to see an article by Gallian on grid-labelling in this section in the near future. Ringel (wrc) has another tree-labelling conjecture: that every tree with $e$ edges, whose nodes are all trivalent or monovalent, can be given a magic labelling: i.e. that the integers $1, 2, \ldots, e$ can be assigned to the edges so that the sum of the three that meet at a node is constant. For example, Figure 1 shows such a labelling for a tree with 29 edges and magic sum 45.

Victor Klee [1970, 63] asked for the maximum length of a $d$-dimensional snake, i.e. a simple cycle without chords in the (edge-skeleton of the) $d$-dimensional cube. Since the work of Evdokimov (1969) it has been known that the order of magnitude is $2^d$: just the constant remains to be determined. Abbott and Katchalski (1988) give a new proof of Evdokimov's theorem.

L. Fejes Tóth [1971, 528] asked for the densest packing of points (or equivalently, of circles of unit diameter) in an infinite strip of width $w$ (or $w + 1$), where any pair of points must be separated by a distance of at least 1. The answer is trivial for $w \leqslant \sqrt{3}/2$, and the intuitively obvious result for $w = n\sqrt{3}/2$ ($n$ an integer) was proved by Molnár (1978). The case $w < \sqrt{2}$ was solved by Kertész (1982) and Füredi (1988) now extends the solution to $w < \sqrt{3}$.

A set of $n$ points in the plane, not all on one (straight) line, determines a set of at least $n$ lines. A line containing just two of the points is an **ordinary line**. If all the lines are ordinary, they are $\binom{n}{2}$ in number, and the points are in **general position**. If all the points, save one, are on a single line, the configuration is a **near-pencil**. Kelly and Moser (1958) proved that at least $3n/7$ of the lines are ordinary. That this is

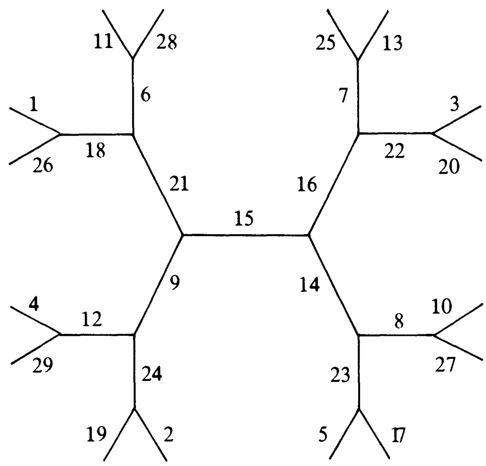



FIG. 1.

best possible can be seen from the **failed Fano configuration**, typified by the vertices, centroid, and midpoints of sides, of a triangle. In Figure 2, the 3 ordinary lines are shown broken, and the 3 ordinary points are circled. Murty [1971, 1000] labelled the points with positive integers, and called a configuration **magic** if the total of the labels on each line was the same. He conjectured that a magic configuration was either in general position, or a near-pencil, or a failed Fano configuration. See also [1988, 464–465 and 932–933].

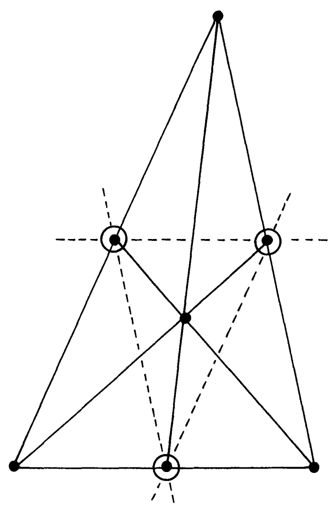


FIG. 2.

A point is *ordinary* if it lies on at least one ordinary line; otherwise it's **special**. The **Sylvester graph** of a configuration is the set of ordinary points and ordinary lines. Smyth (tbp) formulates the conjecture that the Sylvester graph is complete

only for points in general position, or for the failed Fano configuration, and notes that this conjecture and Murty's conjecture both follow from the conjecture

¿ **There are at most 5 special points in a configuration** ?

Molnár [1974, 383] asked for determinants with integer entries, not 0, $\pm 1$, whose value was 1, and remained so when the entries were squared. Several solutions [1975, 999–1000; 1977, 809; 1987, 962] have been published. Rudolf Wytek of the University of Vienna Computer Centre made a search and found six examples, five of which are probably new:

$$\begin{vmatrix} 2 & 3 & 2 \\ 4 & 2 & 3 \\ 9 & 6 & 7 \end{vmatrix} \quad \begin{vmatrix} 2 & 3 & 6 \\ 3 & 2 & 3 \\ 17 & 11 & 16 \end{vmatrix} \quad \begin{vmatrix} 5 & 7 & 6 \\ 6 & 4 & 7 \\ 17 & 16 & 20 \end{vmatrix} \quad \begin{vmatrix} 8 & 7 & 8 \\ 12 & 11 & 7 \\ 17 & 15 & 16 \end{vmatrix} \quad \begin{vmatrix} 10 & 7 & 12 \\ 4 & 2 & 7 \\ 17 & 12 & 20 \end{vmatrix}$$

None of them are special cases of the parametric solutions of Peter Montgomery, Harry Applegate or Morris Newman.

Readers of this section [1987, 963] will know that, modulo the Birch + Swinnerton-Dyer conjecture, Tunnell (1983) has effectively solved the age-old congruent number problem [1980, 43]. Those interested will also wish to read Monsky's (1989) paper, which gives a unified proof that the following numbers are all congruent, where $p_i$ is any prime $\equiv i$, mod 8, and $(p_i|p_j)$ is the Legendre symbol:

$$\begin{aligned} &p_5,\ p_7, 2p_7, 2p_3,\ p_3p_7,\ p_3p_5, 2p_3p_5, 2p_5p_7, \\ &p_1p_5, \text{ provided } (p_1|p_5) = -1, \\ &p_1p_7 \text{ and } 2p_1p_7, \text{ provided } (p_1|p_7) = -1, \\ &2p_1p_3, \text{ provided } (p_1|p_3) = -1. \end{aligned}$$

Most people heeded the advice given on [1983, 35], but an interesting sidelight on the notorious $3x + 1$ problem is thrown by Lagarias (1990), who studies the set of cycles under iteration of all "$3x + k$ functions" where $k$ varies over all positive integers prime to 6, i.e., the set of **rational cycles** of the $3x + 1$ problem. The set of **primitive** such cycles is large and has an interesting structure.

Define a **logarithm** as a one-to-one function, $f$, between the integers $\{1, 2, \dots, k\}$ and the cyclic group of order $k$, such that $f(xy) = f(x) + f(y)$ when $1 \leqslant xy \leqslant k$. More generally, call a subset $S$ of $k$ integers **logarithmic** if there's a group $G$ of order $k$ and a one-to-one map $\phi$: $S \to G$ such that if $x, y, z \in S$ and $z = xy$, then $\phi(z) = \phi(x) + \phi(y)$. Forcade and Pollington (1989) found the counterexamples $k = 195$ and 255 to their [1986, 119] conjecture that $\{1, 2, \dots, k\}$ always have a logarithm. Their conjecture that if $\{1, 2, \dots, k\}$ is logarithmic, then $G$ is commutative, is still open. Chandler (wrc) and Hickerson (wrc) have independently shown that if a noncommutative such $G$ exists, then its centre has index 4, 6 or 8, and Hickerson has also shown that $|G| > 9696$.

For the Mahler-Popken problem [1986, 188], Daniel Rawsthorne (1988) has calculated $f(n)$, the least number of ones needed to represent $n$, using only + and $\times$ (and parentheses) for $n \leqslant 3^{10}$. In this range $f(2^a3^b) = 2a + 3b$, and whenever $f(n) = f(c) + f(d)$ with $c + d = n$, and this minimum is not achieved as a product, then $c$ (or $d$) = 1. He found an additional counterexample, $2^73^2 < 3^55$, with $f(2^73^2) = f(3^55) = 20$, to that found by Selfridge [1987, 965].

In [1987, 967] we outlined a treatment by Lew (tbp) of Ih-Ching Hsu's [1986, 371] functional equation

$$F(x, y) + F(\phi(x), \psi(y)) = F(x, \psi(y)) + F(\phi(x), y).$$

A revision of this is currently being published.

Warren Kokko (1988) answers Corley's [1986, 628] question about how likely are random points in the square to be far apart, but using the "taxicab metric," $|x_i - x_j| + |y_i - y_j|$, in place of the usually Euclidean one.

J. Kincses of Áttila Jozsef University, Szeged, was another who noted that Shapiro's [1987, 46] conjecture on polyhedral cones is a consequence of the Gauss-Bonnet theorem: see Santaló (1976), or, for an elementary proof, McMullen (1975).

There is no lack of interest in diophantine equations [1988, 31]. Jack Good sent a copy of his (1984) paper, summarizing results he obtained in 1962. As he says, "one cannot expect probability arguments... to give accurate results because integers have many hidden properties, but it is interesting to see what they... suggest."

We were just able to slip in a [1988, 32] footnote, earnesting the destruction, by Noam Elkies (1988), of yet another Euler conjecture. It's now well known that he found an infinity of solutions of

$$x^4 + y^4 + z^4 = w^4, \tag{1}$$

the smallest of which is (2682440, 15365639, 18796760, 20615673), and his next smallest involves numbers with 70 decimal digits. Subsequently Roger Frye made a computer search and found the absolutely smallest solution to be (95800, 217519, 414560, 422481): it is presumably not too rash to conjecture that there are infinitely many elliptic curves each of which yields infinitely many solutions of (1). By coincidence, the same week that I heard from Elkies, I received a book which claimed to settle most of the classical unsolved problems of number theory, including a proof that (1) has no nontrivial solutions.

Thomas Berry of Universidad Simon Bolivar, Caracas, Venezuela, observes that the first of equations (9) on [1988, 33] can be written

$$2(s^4 + b^4) + a^4 + c^4 = 2(s^2 + b^2)(a^2 + c^2) \tag{2}$$

(the condition for a point to be at distances $a, b, c$ from the vertices of an isosceles right triangle of sides $s, s\sqrt{2}, s$), and that this, and equation (10) for a point at distances $x$, $y$, $z$ from the vertices of an equilateral triangle of side $s$, both represent **Kummer surfaces**, i.e., quartic surfaces with just 16 singular points. They are not isomorphic, but are of the same special type, known as **tetrahedroids**. There are the following consequences:

(a) A Kummer surface is not rational: there is no general parametric solution of either problem, in the sense that there are no polynomials (resp., rational functions) giving all integer (resp. rational) solutions.

(b) One-parameter families of solutions correspond to parametrizable curves on the surface. For example, the 16 conics (which always exist on a Kummer surface) give, in the equilateral triangle problem, points on the sides and circumcircle. The solution given by Kemnitz [1988, 34] corresponds to the plane section $y + z = 2x$.

(c) The elliptic curves used by Bremner and Guy (1989) form a pencil on (2), and since the surfaces are both tetrahedroids, there may be an elliptic pencil on the "equilateral triangle" surface, which allows an analogous attack.

The problem attributed to Leech [1988, 35], to find distinct triples having the same product and the same sum of squares, goes back at least to Bini (1908), with partial solutions by Dubouis and Mathieu (1909). A neat general solution has been given by John B. Kelly (1990): equal products are related by a matrix of E. T. Bell (1933) and equal sums of squares then give a conic in two-dimensional projective space, on which all rational points can be obtained from a single one. Kelly locates the solutions of Dubouis, of Mathieu, and of Zeitlin (1988) and notes that these last cannot be obtained from those of the first two by specializing the parameters. John Drost (WV), Jean Lagrange (France), Stephane Vandemergel (Belgium) and Mark Weissman (USSR) all sent parametric, though not general, solutions.

There were also several parametric solutions of Mordell's equation (12) [1988, 35]. Zhang Mingzhi (1985) has an earlier publication and Mark Weissman and David Zeitlin sent solutions. Clellie Oursler (wrc) and Judith Longyear (wrc) have submitted more extensive analyses, each of which gives a procedure for finding all solutions. That of Longyear extends to the equation $\Sigma(1/x_i) + \Pi(1/x_i) = 0$ with $n$ ($\geqslant 3$) variables $x_i$ in place of Mordell's 4.

J. A. Macdonald (wrc) of Brighton, Victoria, Australia sends an evaluation of Coxeter's [1988, 330] challenging definite integral in which he notes that "any geometry has been swept under the carpet." Coxeter's own (geometric) solution appeared recently (1989).

Raphael Robinson's [1988, 331] question, how big a slice can you make through a cube? was earlier asked by Hensley (1979) and answered by Keith Ball (1988) who has now (1989) extended the result to $k$-codimensional sections, whose volumes are at most $(\sqrt{2})^k$. S. W. Drury of McGill University, also sends a proof of the original conjecture, and Paul Filliman of Western Washington University, has some conjectures concerning the analogous problem for the regular $n$-dimensional simplex, which he has proved for $1 \leqslant k \leqslant n \leqslant 6$.

Several authors [1988, 428] asked if every latin square of order $n$ has a partial latin transversal (i.e., with all its entries distinct) of size $n - 1$. Their Conjecture 1 [1988, 429], often attributed to Brualdi, is that an equi-$n$-square (in which each symbol occurs exactly $n$ times) has a transversal with at least $n - 1$ distinct symbols. Paul Erdös and Joel Spencer (wrc) have proved the following theorem: an $n \times n$ matrix with various symbols as entries, no symbol appearing more than $k$ times, has a latin transversal if $k \leqslant n/16$.

Jozsef Dénes corrects our editorial remark [1988, 430]: his conjecture is more general than that of Conjecture 1, so that it implies that conjecture, without being equivalent to it. He also reports that although Derienko (1988) appeared to have proved Brualdi's conjecture, the proof is incomplete: attempts to repair it have not been successful.

Gerd Baron, Vienna Technical University, and Johannes Nitsche, University of Minnesota, each use the methods of J. W. Green (1950) to establish Murray Klamkin's [1988, 845] conjectured isoptic characterization of the circle.

Donald Chalice, Bellingham WA, solves Robert J. MacG. Dawson's [1989, 31] paradoxical connections problem by means of an example with an "explosion point" which is a cousin of one of Knaster and Kuratowski (1921).

As a result of correspondence with Stewart Levin, Adam Stinchcombe, Reginald Wood and Clark Carroll himself, and help from Francis Coghlan, we are able to give "all" quartic polynomials which, with their derivatives, have integer roots,

answering the question of Frank Schmidt (1986, 1987) and Clark Carroll [1989, 129]. This can be reduced to finding rational points generated by the point $(-8, -16)$, say, on the rank one elliptic curve $y^2 = x(x+6)(x+24)$. The first few examples have roots $(0, 0, p, q)$, where

$$(p,q) = (308,360),\ (668668,1990440),\ (-277132848044,514660109040),$$
$$(1669580952192186264 0,28041466545675190604).$$

Note that in the third example the roots are of opposite sign. Levin later exhumed a Fermat-type descent given by Allen T. Goldberg in 1976, which reduces any solution to a "smaller" one, culminating in the case of a triple root. A counterexample to Carroll's original conjecture was also found by Michael J. Miller.

Thom Grace and Mark Draughn of Illinois Tech and Charles Ashbacher of Mt. Mercy College sent palindromic answers to Shyam Sunder Gupta's [1989, 425] request for decimal numbers which, with their reversals, give sums and differences which are both squares. They deduce the infinite family of examples

$$\tfrac{1}{2}(2 \times 10^n + 2)^2.$$

Further such families are

$$\tfrac{1}{2}(2 \times 10^{2n} + 2 \times 10^n + 2)^2$$

and

$$\tfrac{1}{2}(2 \times 10^{3n+m} + 2 \times 10^{2n+m} + 2 \times 10^n + 2)^2.$$

Grace and Draughn give all solutions where the sum is $< (2^{16} - 1)^2$. Besides Gupta's examples, there occur the sporadic, nonpalindromic examples

$$281089082 \pm 280980182 = 23708^2, 330^2$$
$$2022652202 \pm 2022562202 = 63602^2, 300^2$$
$$2042832002 \pm 2002382402 = 63602^2, 6360^2$$

It remains to find an infinite family of nonpalindromic examples, and, if possible, to characterize all examples. Similar questions can be asked for bases other than 10.

I am indebted to numerous correspondents, only a few of whom are mentioned; I hope that they, and others, find the section to be a useful clearinghouse for information on unsolved problems.

REFERENCES

H. L. Abbott and M. Katchalski, On the snake in the box problem, *J. Combin. Theory Ser. B.*, **44** (1988) 12–24 [1970, 63].

K. M. Ball, Cube slicing in $\mathbb{R}^n$, *Proc. Amer. Math. Soc.*, 97 (1986) 465–473 [1988, 331].

______, Volumes of sections of cubes and related problems, in Lindenstrauss and Milman (eds), Israeli Seminar on Geometric Aspects of Functional Analysis, 1986/87, *Springer Lecture Notes*, 1317 (1988) 224–231 [1988, 331].

G. Baron, Isoptic characterization of a circle (wrc) [1988, 845].

E. T. Bell, Reciprocal arrays and diophantine analysis, *Amer. J. Math.*, 55 (1933) 50–66 [1988, 35].

U. Bini, Problem 3424. *L'Intermédiaire des Math.*, 15 (1908) 193 [1988, 35].

Andrew Bremner and Richard K. Guy, Delta-lambda configurations in tiling the square, *J. Number Theory* (1989) [1988, 33].

Donald R. Chalice, Explosion sets and paradoxical connexions (wrc) [1989, 31].

K. Chandler, Is a group formed by redefining multiplication on the integers $1, \ldots, n$ necessarily Abelian? preprint, Dalhousie University, 1988 [1986, 119].

H. S. M. Coxeter, Trisecting an orthoscheme, *Computers Math. Appl.*, 17 (1989) 59–71 [1988, 330].
I. I. Derienko, On a conjecture of Brualdi (Russian), *Math. Issladovanie*, (1988) 53–65 [1988, 429].
Dubouis, Mathieu, Réponse 3424, *L'Intermédiaire des Math.*, 16 (1909) 41–42, 112 [1988, 35].
Noam Elkies, On $A^4 + B^4 + C^4 = D^4$, *Math. Comput.*, 51 (1988) 825–835 [1988, 31].
A. A. Evdokimov, Maximal length of a chain in a unit $n$-dimensional cube (Russian), *Mat. Zametki*, 6 (1969) 309–319; MR 40#7945 [1970, 63].
R. W. Forcade and A. D. Pollington, What is special about 195? Groups, $n$th power maps and a problem of Graham, in R. A. Mollin (ed.), Number Theory, Proc. 1st Conf. Canad. Number Theory Assoc., Banff, 1988, de Gruyter, Berlin, New York, 1989.
Zoltán Füredi, The densest packing of equal circles into a parallel strip, *Discrete Comput. Geom.* (1988) [1971, 528].
Joseph A. Gallian, A survey—recent results, conjectures and open problems in labelling graphs, *J. Graph Theory* (tbp) [1969, 1128].
I. J. Good, C202. Probability applied to Diophantine equations, *J. Statist. Comput. Simulation*, 19 no. 4 (1984) 308–313 [1988, 31].
J. W. Green, Sets subtending a constant angle on a circle, *Duke Math J.*, 17 (1950) 263–267 [1988, 845].
Richard K. Guy, Unsolved problems, in R. A. Mollin (ed.) Number Theory, Proc. 1st Conf. Canad. Number Theory Assoc., Banff, 1988, de Gruyter, Berlin, New York, 1989.
D. Hensley, Slicing the cube in $\mathbb{R}^n$ and probability, *Proc. Amer. Math. Soc.*, 73 (1979) 95–100 [1988, 331].
John B. Kelly, Sums of three squares with equal products, this MONTHLY, 97 (1990) [1988, 35].
L. M. Kelly and W. O. J. Moser, On the number of ordinary lines determined by $n$ points, *Canad. J. Math.*, 10 (1958) 210–219 [1971, 1000].
G. Kertész, On a problem of parasites (Hungarian), Dissertation, Budapest, 1982 [1971, 528].
B. Knaster and K. Kuratowski, Sur les ensembles connexes, *Fundamenta Math.*, 2 (1921) 206–255 [1989, 31].
Warren W. Kokko, in *Chance*, 25 (1988), P.O. Box 19818, Denver, CO, 80219.
Jeffery C. Lagarias, The set of rational cycles for the $3x + 1$ problem, *Acta Arith.*, (1990) [1983, 35].
John S. Lew, A functional equation of Ih-Ching Hsu, *Aequationes Math.* (tbp) [1986, 371].
Judith Q. Longyear (wrc), Solving the Mordell equation by classical techniques [1988, 35].
P. McMullen, Non-linear angle-sum relations for polyhedral cones and polytopes, *Math. Proc. Cambridge Philos. Soc.*, 78 (1975) 247–261 [1987, 46].
J. Molnár, Packing of congruent spheres in a strip, *Acta Math. Hungar.*, 31 (1978) 173–183 [1971, 528].
Paul Monsky, Mock Heegner points and congruent numbers, *Math. Z.* (1989) [1980, 43].
Johannes C. C. Nitsche (wrc), Isotopic characterization of a circle.
Clellie C. Oursler (wrc) Mordell's diophantine equation, a search for solutions [1988, 35].
Daniel Rawsthorne, How many ones are needed?, *Fibonacci Quart.*, 27 (1989) [1986, 188].
L. A. Santaló, Integral Geometry and Geometric Probability, Addison-Wesley, Reading, MA, 1976 [1987, 46].
Frank Schmidt, Problems 86-5 and 87-9, *Math. Intelligencer*, 8 No. 2 (1986) 48; 9 No. 3 (1987) 40 [1989, 127].
William F. Smyth (tbp) The number of special points in a Sylvester configuration [1971, 1000].
Jerrold B. Tunnell, A classical diophantine problem and modular forms of weight $3/2$, *Invent. Math.* 72 (1983) 323–334; MR 85d:11046 [1980, 43].
David Zeitlin, Explicit one parameter integer solutions for a coupled pair of diophantine equations, #843-11-50 and #88T-11-147, *Abstracts Amer. Math. Soc.*, 9 No. 3 (Apr. 1988), 239: No. 4 (Aug. 1988), 328 [1988, 35].
Zhang Mingzhi, On the diophantine equation $\frac{1}{x} + \frac{1}{y} + \frac{1}{z} + \frac{1}{w} + \frac{1}{xyzw} = 0$, *Acta Math. Sinica* (N.S.) 1 (1985) 221–224 [1988, 35].

The eigenvalues enable us to tell the sense in which $r$ measures linear association of the points $(x_i, y_i)$ in $R^2$. It is known (compare [**1**, p. 657]) that the quadratic form $Gv \cdot v$, $v$ a unit vector, represents the sum of squares of the distances from the points $(x_i, y_i)$ to the lines through the origin normal to $v$, one for each $v$. The $\lambda_i$ are the extreme values of the form. If $\lambda_2 = 0$, the points are collinear ($\theta = 0$ or $\pi$), falling on the minimal axis of the form. If $\lambda_1 = \lambda_2$ ($\theta = \pi/2$), then $Gv \cdot v$ is constant for all unit vectors. This means that the points are maximally dispersed in the sense that the sum of squares of the distances to *every* line through the origin is the same.

The quotient in (2) thus is 1 at collinearity and zero at maximal dispersion. Thus when $|X| = |Y|$, $r = 0$ signifies maximal dispersion.

The theorem fails when $|X| \neq |Y|$, unless $|r| = 1$. This follows from (1). Also, though $|r|$ is scale invariant, neither the eigenvalues, nor the quotient in (2), nor the configuration of the points in the plane is scale invariant in general.

In the general case of uncorrelated variables ($r = 0$), (1) becomes

$$\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} = \left| \frac{|X|^2 - |Y|^2}{|X|^2 + |Y|^2} \right|.$$

As the ratio of the variances $|X|^2$ and $|Y|^2$ approaches zero (or infinity) the left side approaches 1, and so the points become more and more collinear. Thus uncorrelated variables, after suitable scaling, may give rise to sets of points which are nearly collinear.

There is a duality between an ordered set of $n$ points $(x_i, y_i)$ in the plane and the corresponding vectors $X$ and $Y$ in $n$-space. It is the strength of the linear relationship between $X$ and $Y$ which $r$ measures; we have shown that $r$ is less suitable as a measure of linear association of points, unless $|X| = |Y|$.

**Acknowledgment.** This note was written at the Hebrew University of Jerusalem. Many thanks to Azriel Levy, Chairman of the Mathematics Department for his hospitality during my stay there. Thanks, too, to colleague Matthew Gaffney, for showing me equation (1).

REFERENCE

1. A. Shuchat, Generalized least squares and eigenvalues, this MONTHLY, 92 (1985) 657.

## A Version of Rouché's Theorem for Continuous Functions

A. TSARPALIAS
*Department of Mathematics, University of Athens, 157 81 Athens, Greece*

Let $D$ denote the closed unit disc in $\mathbb{C}$. We prove the following version of Rouché's theorem, assuming the results and using the notation of [**1**].

THEOREM. *Let $f: D \to \mathbb{C}$ be a continuous function and $g$ an analytic function on a neighborhood of $D$. If $|f(\zeta)| \le |g(\zeta)|$ for every $\zeta \in \partial D$ and $g$ has a root in $D$, then $f + g$ has a root in $D$.*

*Proof.* We assume that $g(\zeta) \neq 0$ for every $\zeta \in \partial D$, the result otherwise being obvious.

The eigenvalues enable us to tell the sense in which $r$ measures linear association of the points $(x_i, y_i)$ in $R^2$. It is known (compare [1, p. 657]) that the quadratic form $Gv \cdot v$, $v$ a unit vector, represents the sum of squares of the distances from the points $(x_i, y_i)$ to the lines through the origin normal to $v$, one for each $v$. The $\lambda_i$ are the extreme values of the form. If $\lambda_2 = 0$, the points are collinear ($\theta = 0$ or $\pi$), falling on the minimal axis of the form. If $\lambda_1 = \lambda_2$ ($\theta = \pi/2$), then $Gv \cdot v$ is constant for all unit vectors. This means that the points are maximally dispersed in the sense that the sum of squares of the distances to *every* line through the origin is the same.

The quotient in (2) thus is 1 at collinearity and zero at maximal dispersion. Thus when $|X| = |Y|$, $r = 0$ signifies maximal dispersion.

The theorem fails when $|X| \neq |Y|$, unless $|r| = 1$. This follows from (1). Also, though $|r|$ is scale invariant, neither the eigenvalues, nor the quotient in (2), nor the configuration of the points in the plane is scale invariant in general.

In the general case of uncorrelated variables ($r = 0$), (1) becomes

$$\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} = \left| \frac{|X|^2 - |Y|^2}{|X|^2 + |Y|^2} \right|.$$

As the ratio of the variances $|X|^2$ and $|Y|^2$ approaches zero (or infinity) the left side approaches 1, and so the points become more and more collinear. Thus uncorrelated variables, after suitable scaling, may give rise to sets of points which are nearly collinear.

There is a duality between an ordered set of $n$ points $(x_i, y_i)$ in the plane and the corresponding vectors $X$ and $Y$ in $n$-space. It is the strength of the linear relationship between $X$ and $Y$ which $r$ measures; we have shown that $r$ is less suitable as a measure of linear association of points, unless $|X| = |Y|$.

**Acknowledgment.** This note was written at the Hebrew University of Jerusalem. Many thanks to Azriel Levy, Chairman of the Mathematics Department for his hospitality during my stay there. Thanks, too, to colleague Matthew Gaffney, for showing me equation (1).

REFERENCE

1. A. Shuchat, Generalized least squares and eigenvalues, this MONTHLY, 92 (1985) 657.

## A Version of Rouché's Theorem for Continuous Functions

A. TSARPALIAS

*Department of Mathematics, University of Athens, 157 81 Athens, Greece*

Let $D$ denote the closed unit disc in $\mathbb{C}$. We prove the following version of Rouché's theorem, assuming the results and using the notation of [1].

THEOREM. *Let $f: D \to \mathbb{C}$ be a continuous function and $g$ an analytic function on a neighborhood of $D$. If $|f(\zeta)| \leq |g(\zeta)|$ for every $\zeta \in \partial D$ and $g$ has a root in $D$, then $f + g$ has a root in $D$.*

*Proof.* We assume that $g(\zeta) \neq 0$ for every $\zeta \in \partial D$, the result otherwise being obvious.

One can easily find some other consequences. We mention, for instance, the following well-known fact which contains the fundamental theorem of algebra: Let $f\colon \mathbb{C} \to \mathbb{C}$ be a continuous function such that there exists a positive integer $m$ and a complex number $c \neq 0$ with $\lim_{z\to\infty} z^{-m}f(z) = c$. Then $f$ has a root.

REFERENCES

1. J. B. Conway, Functions of One Complex Variable, Springer-Verlag, New York, 1978.
2. J. Dugundji, Topology, Allyn and Bacon, Boston, 1966.

# Another Example of an Exotic Function

Victor Pambuccian
*Mountain House School, 12 Lake Placid Club Drive, Lake Placid, NY 12946*

The purpose of this note is to exhibit a function, which is zero almost everywhere (i.e., a function that differs from zero on a *meager* set; here "meager" stands for "of Lebesgue measure zero and of first category") and takes every real value in any given interval. The construction is different from those in [3, Ch. 8, ex. 27], [**1**, Ch. I, ex. 1.2 and Th. 3.4], [**4**], and [**5**].

Let $G$ be any meager additive subgroup of $\mathbb{R}$ of cardinality $2^{\aleph_0}$. Then $\mathbb{Q}\cdot G = \{qg \mid q \in \mathbb{Q},\ g \in G\}$ is meager too. If $a \in \mathbb{R}\setminus\mathbb{Q}\cdot G$ and $H = a^{-1}G$, then $H \cap \mathbb{Q} = \{0\}$, $|H| = 2^{\aleph_0}$ and $H$ is meager. Hence every $x$ in $A = H + \mathbb{Q}$ can be uniquely written as $x = h + q$, with $h$ in $H$, $q$ in $\mathbb{Q}$. Let $\phi\colon H \to \mathbb{R}$ be any one-to-one function and define $\psi$ by

$$\psi(x) = \begin{cases} \phi(h), & \text{if } x \in A,\ x = h + q \\ 0, & \text{otherwise.} \end{cases}$$

Then $\psi$ has all the required properties. The construction depends on the existence of a meager additive subgroup of $\mathbb{R}$ of cardinality $2^{\aleph_0}$. The additive subgroup of $\mathbb{R}$ generated by the set

$$\left\{ x \in \mathbb{R} \,\middle|\, x = \sum_{i=1}^{\infty} \frac{a_i}{(2i)!}, \qquad 0 \le a_i < 2i \right\}$$

is such an example (for details concerning its measurability see [**2**, p. 191]).

Any example of such an exotic function can be used to show the ubiquity of everywhere discontinuous functions that have the Darboux property, in the following sense (see [**1**, Th. 3.4] and [**5**]; the result was first proved in [**4**]): For any function $f\colon \mathbb{R} \to \mathbb{R}$, there is a function $g\colon \mathbb{R} \to \mathbb{R}$,

$$g(x) = \begin{cases} f(x), & \text{if } x \in \mathbb{R}\setminus A \\ \psi(x), & \text{if } x \in A \end{cases}$$

that differs from $f$ on a set of measure zero and of first category and has the intermediate value property, although it is nowhere continuous.

REFERENCES

1. A. M. Bruckner, Differentiability of real functions, Lecture Notes in Mathematics, Vol. 659, Springer-Verlag, Berlin, Heidelberg, New York, 1978.

One can easily find some other consequences. We mention, for instance, the following well-known fact which contains the fundamental theorem of algebra: Let $f\colon \mathbb{C} \to \mathbb{C}$ be a continuous function such that there exists a positive integer $m$ and a complex number $c \neq 0$ with $\lim_{z\to\infty} z^{-m}f(z) = c$. Then $f$ has a root.


REFERENCES

1. J. B. Conway, Functions of One Complex Variable, Springer-Verlag, New York, 1978.
2. J. Dugundji, Topology, Allyn and Bacon, Boston, 1966.


# Another Example of an Exotic Function


VICTOR PAMBUCCIAN

*Mountain House School, 12 Lake Placid Club Drive, Lake Placid, NY 12946*


The purpose of this note is to exhibit a function, which is zero almost everywhere (i.e., a function that differs from zero on a *meager* set; here "meager" stands for "of Lebesgue measure zero and of first category") and takes every real value in any given interval. The construction is different from those in [3, Ch. 8, ex. 27], [**1**, Ch. I, ex. 1.2 and Th. 3.4], [**4**], and [**5**].

Let $G$ be any meager additive subgroup of $\mathbb{R}$ of cardinality $2^{\aleph_0}$. Then $\mathbb{Q}\cdot G = \{qg \mid q \in \mathbb{Q},\ g \in G\}$ is meager too. If $a \in \mathbb{R}\setminus\mathbb{Q}\cdot G$ and $H = a^{-1}G$, then $H \cap \mathbb{Q} = \{0\}$, $|H| = 2^{\aleph_0}$ and $H$ is meager. Hence every $x$ in $A = H + \mathbb{Q}$ can be uniquely written as $x = h + q$, with $h$ in $H$, $q$ in $\mathbb{Q}$. Let $\phi\colon H \to \mathbb{R}$ be any one-to-one function and define $\psi$ by

$$\psi(x) = \begin{cases} \phi(h), & \text{if } x \in A,\ x = h + q \\ 0, & \text{otherwise.} \end{cases}$$

Then $\psi$ has all the required properties. The construction depends on the existence of a meager additive subgroup of $\mathbb{R}$ of cardinality $2^{\aleph_0}$. The additive subgroup of $\mathbb{R}$ generated by the set

$$\left\{ x \in \mathbb{R} \,\middle|\, x = \sum_{i=1}^{\infty} \frac{a_i}{(2i)!}, \qquad 0 \le a_i < 2i \right\}$$

is such an example (for details concerning its measurability see [**2**, p. 191]).

Any example of such an exotic function can be used to show the ubiquity of everywhere discontinuous functions that have the Darboux property, in the following sense (see [**1**, Th. 3.4] and [**5**]; the result was first proved in [**4**]): For any function $f\colon \mathbb{R} \to \mathbb{R}$, there is a function $g\colon \mathbb{R} \to \mathbb{R}$,

$$g(x) = \begin{cases} f(x), & \text{if } x \in \mathbb{R}\setminus A \\ \psi(x), & \text{if } x \in A \end{cases}$$

that differs from $f$ on a set of measure zero and of first category and has the intermediate value property, although it is nowhere continuous.


REFERENCES

1. A. M. Bruckner, Differentiability of real functions, Lecture Notes in Mathematics, Vol. 659, Springer-Verlag, Berlin, Heidelberg, New York, 1978.

2. P. Erdös, K. Kunen, and R. Daniel Mauldin, Some additive properties of sets of real numbers, *Fund. Math.*, 113 (1981) 187–199.
3. B. R. Gelbaum and J. M. H. Olmsted, Counterexamples in Analysis, Holden-Day, San Francisco, 1964.
4. S. Marcus, Sur une propriété appartenant à toutes les fonctions réelles d'une variable réelle, *Indian J. Math*,. 9 (1967) 457–460.
5. Solution to Problem 6505, this MONTHLY, 94 (1987) 560.

EDITED BY MELVIN HENRIKSEN AND STAN WAGON

# A Fresh(man) Treatment of Determinants

KENNETH P. BOGART
*Department of Mathematics, Dartmouth College, Hanover, NH 03755*

As a result of the inclusion of matrix algebra in our department's freshman-level discrete mathematics course, I have had to grapple with how to teach determinants to freshmen. Since the goal of a discrete mathematics course is to teach students to *understand* mathematics, it is essential to have a conceptual definition of a determinant. Since induction and recursion are central concepts in a discrete mathematics course, it is important to include the recursive row-expansion formula for a determinant. Since linearity is *not* a central concept of the discrete mathematics course, the description of a determinant as an $n$-linear function of a special kind is not an especially desirable conceptual definition, contrary to its desirability in a linear algebra course.

Matrix algebra topics one is likely to introduce in a discrete mathematics course include row-reduction and inverse matrices. This makes it seem natural to students that we search for a single number, called the determinant, associated with a matrix in such a way that it is nonzero if and only if the matrix is invertible. It also makes it reasonable to try to describe such a determinant by the elementary row operations. The theme of this paper is that the most appropriate elementary conceptual approach to defining determinants is the following:

- Describe how a determinant function should interact with the elementary row operations.
- Observe that, if the determinant function described exists, then the value is uniquely determined by the description.
- Finally, exhibit a formula that demonstrates there is a function satisfying the description.

The theme is not original with this paper; it is certainly part of the folklore of determinants and is the approach taken, for example, by Kumpel and Thorpe **[3]** (where proofs are restricted to the $2 \times 2$ case)[1]. In order to give a quick treatment for the $n$ by $n$ case, it is tempting to give a known formula for the determinant, such as the permutation or the column expansion formula, derive the linear and alternating properties from it, and derive the row-operation description from these. (In fact, we can even skip over the intermediate step of linearity for a time savings.) However, this approach leaves students wondering how they were supposed to guess the formula. There is a similar problem when one begins with the more standard description of the determinant as a multilinear alternating function and similarly pulls a formula from "thin air" to show determinants exist. In Apostol's elegant and elementary treatment of determinants as multilinear functions **[1]**, he overcomes this

[1] See also the beautiful book by Samelson **[4]** which is, unfortunately, out of print.

problem by *deriving* the column expansion formula from the usual axioms, developing many useful results about determinants along the way. While this is my preferred approach for sophisticated students or for a linear algebra course, students whose mathematical skills are more mechanical than conceptual have difficulty with it. In this paper I present a more mechanical approach to deriving the column-expansion formula from row-reduction axioms and verifying that the formula satisfies the axioms. While this treatment is similar to Apostol's in that both make essential use of the basic facts about determinants of products and transposes, the present treatment is quite different from and more mechanical than Apostol's. This paper merely outlines the approach; the reader may find the details in [2].

**The row operations.** We give a slightly unusual description of the elementary row operations. We use

$$\begin{bmatrix} -R1- \\ -R2- \\ \vdots \\ -Rn- \end{bmatrix}$$

to stand for the matrix with rows $R1$ through $Rn$; we avoid subscripts so that $Ri_j$ stands for the $j$th element of row $i$ while $R_{ij}$ stands for the entry in row $i$ and column $j$ of a rectangular matrix $R$. The row operations we use are:

(1) Factor an $r$ from row, $i$. This replaces

$$\begin{bmatrix} -R1- \\ \vdots \\ -rRi- \\ \vdots \\ -Rn- \end{bmatrix} \text{ by } \begin{bmatrix} -R1- \\ \vdots \\ -Ri- \\ \vdots \\ -Rn- \end{bmatrix}.$$

(2) Add $r$ times row $i$ to row $j$.

(3) Interchange row $i$ and row $j$.

**Defining determinants axiomatically by the effect of row operations.** As usual, we may use row-reduction to discover that the $2 \times 2$ matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ has an inverse if and only if $ad - bc \neq 0$, so we define the determinant of this matrix to be $ad - bc$.

By examining how the three kinds of row operations affect the value of $ad - bc$ we decide on the following axioms describing how the determinant interacts with row operations.

(1) If $A'$ is obtained from $A$ by factoring out $r$ from row $i$, then $\det A = r \det A'$. (We allow $r$ to be 0.) More symbolically

$$\det \begin{bmatrix} -R1- \\ \vdots \\ -rRi- \\ \vdots \\ -Rn- \end{bmatrix} = r \det \begin{bmatrix} -R1- \\ \vdots \\ -Ri- \\ \vdots \\ -Rn- \end{bmatrix}.$$

(2) If $A'$ is obtained from $A$ by adding a multiple of row $j$ to row $i$, then $\det A' = \det A$. More symbolically

$$\det\begin{bmatrix} -R1- \\ \vdots \\ Ri + rRj \\ \vdots \\ -Rj- \\ \vdots \\ -Rn- \end{bmatrix} = \det\begin{bmatrix} -R1- \\ \vdots \\ -Ri- \\ \vdots \\ -Rj- \\ \vdots \\ -Rn- \end{bmatrix}.$$

(3) If $A'$ is obtained from $A$ by interchanging row $i$ and row $j$, then $\det A' = -\det A$. More symbolically

$$\det\begin{bmatrix} -R1- \\ \vdots \\ -Ri- \\ \vdots \\ -Rj- \\ \vdots \\ -Rn- \end{bmatrix} = -\det\begin{bmatrix} -R1- \\ \vdots \\ -Rj- \\ \vdots \\ -Ri- \\ \vdots \\ -Rn- \end{bmatrix}.$$

From Axiom 1 with $r = 0$ we see that if a matrix has a row of zeros then its determinant is 0; from Axioms 1 and 2 together we see if a matrix may be row-reduced to a matrix with a row of zeros, then its determinant is 0. Further we see that Axiom 3 may be derived from Axioms 1 and 2. We also see that we can reduce the problem of computing any other determinant to the problem of computing the determinant of an identity matrix. In analogy with the case of $2 \times 2$ matrices we adopt one more axiom.

(4) If $I$ is an identity matrix, then $\det I = 1$.

**Computing determinants if they exist.** From Axioms 1–4 it is straightforward to prove that *if there is* a determinant function defined on $n \times n$ matrices, then

(1) $\det A = 0$ if and only if $A$ is not invertible.
(2) If $A$ is invertible, then $\det A$ is the product of the *elementary factors* that arise from a row-reduction of $A$ to the identity: There is a factor of $r$ for each time we factor $r$ from row $j$, a factor of $-1$ for each time we interchange row $i$ and row $j$, and a factor of 1 for each time we add a multiple of row $j$ to row $i$.

As a consequence, if there *is* a determinant function, then there is *only* one determinant function, the one computed by this method.

Students can understand (and prove) the facts about determinants we've discussed so far as easily as they can understand (and prove) facts about the computation of inverses by row-reduction. The hard part is to *prove* that a determinant function really does exist. For this purpose, we have to figure out a

formula for the determinant and then prove it satisfies rules 1–4. A formula that works for this purpose is the formula we know as the cofactor expansion on the first column. However, there are two stumbling blocks. First, in order to have an intellectually satisfying approach, we should *derive* the formula from what we already know. Second, we have to show that the formula satisfies our axioms.

**The additive property.** The formula for expansion on row $i$ follows almost trivially from the $n$-linearity point of view. With the point of view adopted here it is necessary to prove the formula

$$\det\begin{bmatrix} -R1- \\ \vdots \\ -Ri- \\ \vdots \\ Rj + R'j \\ \vdots \\ -Rn- \end{bmatrix} = \det\begin{bmatrix} -R1- \\ \vdots \\ -Ri- \\ \vdots \\ -Rj- \\ \vdots \\ -Rn- \end{bmatrix} + \det\begin{bmatrix} -R1- \\ \vdots \\ -Ri- \\ \vdots \\ -R'j- \\ \vdots \\ -Rn- \end{bmatrix} \tag{5}$$

in order to determine that

$$\det\begin{bmatrix} a_{11}a_{12} \cdots a_{1n} \\ -R2- \\ -R3- \\ \vdots \\ -Rn- \end{bmatrix} = \det\begin{bmatrix} a_{11}0\text{–}0 \\ -R2- \\ \vdots \\ -Rn- \end{bmatrix} + \det\begin{bmatrix} 0\,a_{12}0\text{–}0 \\ -R_2- \\ \vdots \\ -Rn- \end{bmatrix} + \cdots + \det\begin{bmatrix} 0\text{–}0\,a_{1n} \\ -R2- \\ \vdots \\ -Rn- \end{bmatrix}.$$

The row expansion formula follows because

$$\det\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & a_{11} & a_{12} & \cdots & a_{1,n-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & a_{n-1,1} & a_{n-1,2} & \cdots & a_{n-1,n-1} \end{bmatrix}$$

satisfies all the properties of an $(n-1) \times (n-1)$ determinant and thus if determinants exist, the determinant of this matrix must be the determinant of the $(n-1) \times (n-1)$ matrix in the lower right-hand corner. It is possible to derive (5) from (1) and (2) as follows. If the matrix with rows $R1$ through $Rn$ can be row reduced to the identity, then by first reducing it to the identity and then combining rows of the identity we get

$$R'j = a_1R1 + a_2R2 + \cdots + a_nRn.$$

Applying rules (1) and (2) to the left-hand and right-hand determinants in (5) converts $R'j$ to $a_jRj$, and so in this case (5) follows. Similar computations (using a linear combination of the rows that is zero) let us prove (5) in the case where the matrix with rows $R1$ through $Rn$ is not invertible. This computation (which demonstrates linearity) is the cost of not assuming linearity.

**Determinants of transposes.** Now how do we derive the column expansion formula from the row expansion formula? Of course we want to use the fact that the determinant of a matrix equals the determinant of its transpose. A matrix is a product of elementary matrices (matrices obtained by acting on $I$ with a row operation) if and only if it is invertible. The determinant of an elementary matrix is the elementary factor of the operation used to reduce it back to the identity. Since the transpose of an elementary matrix is elementary (and has the same elementary factor) and the transpose of a product is the product of the transposes in the reverse order, this proves that the determinant of a matrix and its transpose are equal (notice a matrix is *not* a product of elementary matrices if and only if its transpose is not either).

**Proving determinants exist.** Now we have everything needed to derive the column expansion formula, completing the derivation of a valid formula for the determinant function if such a function exists. Now, to show that a determinant function *does* exist, we show that the column expansion formula satisfies axioms 1–4. Axioms 1 and 4 are easy; Axiom 3 is unnecessary because it follows from 1 and 2. In the same way, Axiom 2 may be derived from Axiom 2′ below and Axiom 1.

2′). If $A'$ is obtained from $A$ by adding row $i$ to row $i+1$ or row $i-1$ then $\det A' = \det A$.

THEOREM. *The formula* $\det A = \sum_{k=1}^{n}(-1)^{k+1}a_{k1}\det A(k,1)$ *for expansion of a determinant along the first column (in which* $A(i,j)$ *stands for the matrix obtained from* $A$ *by deleting row* $i$ *and column* $j$*) satisfies axiom* 2′.

*Proof.* We consider adding row $i$ to row $i+1$. We assume inductively that the function defined by the expansion formula satisfies axiom 2′ for $(n-1)\times(n-1)$ determinants.

We use $A'$ to denote the matrix obtained from $A$ by adding row $i$ to row $i+1$. If $k\neq i$ and $k\neq i+1$, then $a'_{k1}\det A'(k,1) = a_{k1}\det A(k,1)$ by the induction hypothesis applied to $A(k,1)$. Further, using $\overline{R}$ to stand for row $R$ with the first entry deleted, we may write

$$(-1)^{i+1}a'_{i1}\det A'(i,1) + (-1)^{i+2}a'_{i+1,1}\det A'(i+1,1)$$

$$= (-1)^{i+1}a_{i1}\det\begin{bmatrix}\overline{R}1\\ \overline{R}2\\ \overline{R}i-1\\ \overline{R}(i+1)+\overline{R}i\\ \vdots\\ \overline{R}n\end{bmatrix} + (-1)^{i+2}(a_{i1}+a_{i+1,1})\det A(i+1,1)$$

$$= (-1)^{i+1}a_{i1}(\det A(i,1) + \det A(i+1,1)) + (-1)^{i+2}a_{i1}\det A(i+1,1)$$

$$+(-1)^{i+2}a_{i+1,1}\det A(i+1,1)$$

$$= (-1)^{i+1}a_{i1}\det A(i,1) + (-1)^{(i+1)+1}a_{i+1,1}\det A(i+1,1),$$

which shows that the overall contribution of rows $i$ and $i + 1$ of $A'$ to the column expansion formula is identical to the overall contribution of rows $i$ and $i + 1$ of $A$ to the column expansion formula. Thus the column expansion formula gives exactly the same results for $A$ and $A'$. □

The theorem was stated and proved formally because the computation is easier to show in detail than to describe loosely but accurately. On the basis of this theorem, we may conclude that the column expansion formula does define a determinant function. Since the proof of the theorem is easier than a proof that the column expansion formula is $n$-linear and alternating, we get a payoff here for basing our determinant axioms on row-reduction operations rather than $n$-linearity.

**A final note.** One disadvantage of this approach to determinants is that we must derive the additivity property of determinants as a theorem. A second disadvantage is that as we prove our results (as in the proof of additivity) we must often deal with noninvertible matrices separately from invertible ones.

Balancing this, there are three advantages. First, we have a simple conceptual definition of determinants requiring only modest effort to prove that determinants exist.

Second, the row-reduction determinant axioms relate intimately to the volume of the parallelepiped (in $n$ dimensions) spanned by the rows of a matrix. Multiplying a row by a number $r$ multiplies the volume by $|r|$. Adding one row to another changes the parallelogram spanned by those two rows to another one with the same area, so it changes neither the area of a 2-dimensional base nor the "$n$ − 2-dimensional height" and thus does not change the volume. (Notice by Axiom 2′ that we don't need to worry about adding any other multiple of row $i$ to row $j$.) Interchanging two rows doesn't change the parallelepiped. The volume of the parallelepiped spanned by an identity matrix could only be 1. Thus the volume is the product of the absolute values of the elementary factors of the determinant, so it is the absolute value of the determinant. We thus see that in 3-space the volume of a parallelpiped is the absolute value of the determinant of the associated matrix, and in higher-dimensional spaces, determinants are a natural tool for *defining* volume.

Third, the computations, though there are many, are not difficult. The most complicated result conceptually appears to be the proof of the addition formula; the most detailed computation needed in a proof is the proof we gave that the column expansion formula obeys rule 2′. After a brief treatment of elementary matrices as a part of studying invertibility, I find that I (and others with many different levels of experience) can treat the outline above in three (or perhaps two) class periods at Dartmouth. The students (mostly freshmen) learn what we want them to, namely they can compute determinants either by an expansion formula or by row-reduction, they seem to have an idea of what the determinant is and does, and they realize that defining something axiomatically doesn't necessarily mean that it exists.

REFERENCES

1. Tom M. Apostol, Calculus, Volume II, second ed., Wiley, New York, 1969.
2. Kenneth P. Bogart, Discrete Mathematics, D. C. Heath and Co., Lexington, Mass., 1988.
3. Paul G. Kumpel and John A. Thorpe, Elementary Linear Algebra, Saunders College Publishing, Philadelphia, 1983.
4. Hans Samelson, An Introduction to Linear Algebra, Wiley, New York, 1974.

# On the Vandermonde Matrix

JOSEPH J. RUSHANAN*
*The MITRE Corporation, M/S E025, Bedford, MA 01730*

The Vandermonde matrix is ubiquitous in mathematics and engineering. Its uses include polynomial interpolation [**1**, **4**], coding theory [**2**, **5**], and signal processing, where the matrix for a discrete Fourier transform is a Vandermonde matrix. There is an extensive literature on numerically solving systems of linear equations when the matrix is given by a Vandermonde matrix [**1**, **3**]. And the Vandermonde determinant is especially common in the classroom, where several standard methods are used to give its elegant formula (see, e.g., [**1**, **2**, **5**]). In these examples, specific properties of the Vandermonde matrix are usually developed as needed, using isolated, albeit interesting, techniques.

The purpose of this note is to present a unified approach to the Vandermonde matrix. Hence, our goal is to derive properties of the Vandermonde matrix not in a faster manner, but in a more integrated and straightforward one. This approach will yield the inverse and determinant of a Vandermonde matrix, along with results on the factorization of a Vandermonde (including its LU-decomposition). All of these properties are known in one form or another; see the references cited above.

Besides integrating all of these properties of the Vandermonde matrix into one discussion, our goal is also to offer a bridge between two disciplines of thought: on the one hand that of the engineer, speaking of transforms, and on the other that of the algebraist, speaking of homomorphisms. Perhaps in this way students with a "purer" background can easily make a transition to more "applied" subjects, and vice versa.

The main technique is to associate a given Vandermonde matrix with a certain linear transformation. Let $\alpha_0, \alpha_1, \dots, \alpha_{n-1}$ be distinct elements from a field $\mathbb{F}$ and define the polynomial

$$f(x) = (x - \alpha_0)(x - \alpha_1) \cdots (x - \alpha_{n-1}).$$

Let $R$ be the ring of polynomials over $\mathbb{F}$ modulo $f(x)$, represented by all polynomials over $\mathbb{F}$ of degree less than $n$. We define a linear transformation $\varphi$ on $R$ to $\mathbb{F}^n$ by

$$\begin{aligned} &\varphi\colon R \to \mathbb{F} \times \cdots \times \mathbb{F} \\ &\varphi(g(x)) = (g(\alpha_0), \dots, g(\alpha_{n-1})), \quad \text{where } g(x) \in R. \end{aligned} \tag{1}$$

In engineering parlance, $\varphi$ is a transform, and $\varphi(g(x))$ is called the transform of $g(x)$. The fact that the transform of a sum is the sum of the transforms means that $\varphi$ is an additive homomorphism. The fact that $\varphi$ changes polynomial multiplication (convolution "in the time domain") to pointwise multiplication ("in the frequency domain") means that $\varphi$ is a multiplicative homomorphism. Thus $\varphi$ is also a ring homomorphism.

It is easy to see that $\varphi$ is in fact an isomorphism. Notice that $\varphi$ is one-to-one, since if the transform of $g(x)$ is the zero vector, then $x - \alpha_i$ divides $g(x)$ for $i =$

*The author thanks the referee for suggestions leading to the improvement of this paper. This work was supported by the MITRE Sponsored Research Program.

$0, 1, \ldots, n-1$ and so $g(x)$ is congruent to zero modulo $f(x)$. But then $\varphi$ must also be onto, since it is an injection from an $n$-dimensional space to an $n$-dimensional space.

As an example, consider the discrete Fourier transform over the complex numbers, so that $\mathbb{F} = \mathbb{C}$. Let $\omega$ be a primitive $n$th root of unity and set $\alpha_i = \omega^i$, for $i = 0, 1, \ldots, n-1$. Then $f(x) = x^n - 1$. It should be noted that discrete Fourier transforms have also been used over finite fields; see [**2**].

Suppose we choose the usual basis $\{1, x, \ldots, x^{n-1}\}$ for $R$. Then the matrix for $\varphi$ under this basis is the Vandermonde matrix

$$V = \begin{pmatrix} 1 & 1 & & 1 \\ \alpha_0 & \alpha_1 & & \alpha_{n-1} \\ \alpha_0^2 & \alpha_1^2 & & \alpha_{n-1}^2 \\ \vdots & \vdots & \cdots & \vdots \\ \alpha_0^{n-1} & \alpha_1^{n-1} & & \alpha_{n-1}^{n-1} \end{pmatrix}. \tag{2}$$

In the above example of the discrete Fourier transform, the $ij$ entry of $V$ is $\omega^{ij}$, where $i$ and $j$ vary from 0 to $n-1$.

Our goal is to exploit the transform presented in (1) to garner results about $V$. In order to do this, we always identify polynomials of degree up to $n-1$ over $\mathbb{F}$ with row vectors in $\mathbb{F}^n$ by taking the coefficients in order by increasing degree. This order is forced by the ordering of the basis used to define $V$ in (2). Thus, if $g(x)$ is identified with the row vector $\mathbf{g}$, then $\varphi(g(x)) = \mathbf{g}V$.

We first find the inverse of $V$, i.e., a matrix $W$ satisfying $WV = I$. If $\mathbf{w}_i$ is the $i$th row of $W$, then $\mathbf{w}_i V$ has a 1 in the $i$th position and zeros everywhere else. Let $w_i(x)$ be the polynomial corresponding to $\mathbf{w}_i$. Then $w_i(\alpha_j)$ is 1 if $i = j$ and 0 otherwise. Thus, $w_i(x)$ must be divisible by $x - \alpha_j$ for $i \neq j$. This leads to the definitions

$$w_i(x) = \frac{\prod_{j \neq i} (x - \alpha_j)}{\prod_{j \neq i} (\alpha_i - \alpha_j)}, \qquad \text{for } i = 0, 1, \ldots, n-1,$$

where we have normalized to get $w_i(\alpha_i) = 1$. Hence the inverse of $V$ is the matrix $W$ whose rows are the vectors $\mathbf{w}_i$ generated by the $w_i(x)$. This technique is a straightforward application of Lagrange interpolation for polynomials (see, e.g., [**1, 4**]).

Suppose next that $V$ factors into two $n \times n$ matrices, say $V = V_1 V_2$; necessarily $V_1$ and $V_2$ are nonsingular. Thus, we have $V_1^{-1} V = V_2$, which is reminiscent of the equation $WV = I$ that we used to find the inverse of $V$. Associate polynomials in $x$ with the rows of $V_1^{-1}$. Notationally,

$$V_1^{-1} = \begin{pmatrix} m_0(x) \\ m_1(x) \\ \vdots \\ m_{n-1}(x) \end{pmatrix}. \tag{3}$$

REFERENCES

1. Richard Bellman, Introduction to Matrix Analysis, 2nd ed., McGraw-Hill, New York, 1970.
2. Richard E. Blahut, Theory and Practice of Error Control Codes, Addison Wesley, Reading, Mass., 1983.
3. Gene H. Golub and Charles F. Van Loan, Matrix Computation, Johns Hopkins Univ. Press, Baltimore, 1983, pp. 119–124.
4. Allen Klinger, The Vandermonde matrix, this MONTHLY, 74 (1967) 571–574.
5. Vera Pless, Introduction to the Theory of Error-Correcting Codes, John Wiley, New York, 1982.

# Rings With Invertible Regular Elements

F. C. LEARY
*Department of Mathematics, St. Bonaventure University, St. Bonaventure, NY 14778*

Let $R$ be a ring (assumed to have 1 and associative multiplication). Beginners in abstract algebra often find it annoying that $R$ may have the very unintegerlike property of having *zero divisors*, i.e., nonzero elements $x$ and $y$ for which $xy = 0$ (e.g., in $\mathbf{Z}/6\mathbf{Z}$, the ring of integers mod 6, $2 \cdot 3 = 0$). If $R$ is noncommutative, we call $x$ a *left zero divisor* and $y$ a *right zero divisor*. For commutative $R$, no such distinction is necessary.

Nonzero elements that are neither left nor right zero divisors are called *regular*. Regular elements resemble units (i.e., invertible elements) in that they may be cancelled. Of course, regular elements need not be units, e.g., the element 2 in $\mathbf{Z}$, the ring of integers. On the other hand, there are rings in which all regular elements are units, e.g., any division ring.

This note was motivated by class discussion relating to the following problem, which appears in [**3**, Chap. 20, Exer. C1]: Let $R$ be a finite commutative ring. Then every nonzero element of $R$ is either a zero divisor or a unit. The discussion centered around the fact that such an $R$ must then have the property $(P)$: all regular elements are units.

At this point, two questions naturally arise.

1. Do all finite rings have property $(P)$?
2. Is there a way to distinguish the rings that have property $(P)$ from those that do not?

The answer to (1) is yes and the proof is revealing.

First let $R$ be commutative and $x$ in $R$ be regular. Then left multiplication by $x$ provides an injective function $f: R \to R$. Since $R$ is a finite set, $f$ must also be surjective. Thus, $f(a) = xa = 1$ for some $a$ in $R$ and $x$ is a unit. This same argument applied to both left and right multiplication by $x$ shows that any finite ring has property $(P)$.

The key ingredient in the preceding discussion is the fact that for finite sets $S$, every injective $f: S \to S$ is necessarily surjective. A similar requirement on module homomorphisms is closely related to our discussion (see Theorem 1) and allows a characterization of commutative rings having property $(P)$.

Let $R$ be commutative and consider $R$ as an $R$-module. Any $R$-module homomorphism $f: R \to R$ is multiplication by an element $x$ of $R$. It is immediate that

REFERENCES

1. Richard Bellman, Introduction to Matrix Analysis, 2nd ed., McGraw-Hill, New York, 1970.
2. Richard E. Blahut, Theory and Practice of Error Control Codes, Addison Wesley, Reading, Mass., 1983.
3. Gene H. Golub and Charles F. Van Loan, Matrix Computation, Johns Hopkins Univ. Press, Baltimore, 1983, pp. 119–124.
4. Allen Klinger, The Vandermonde matrix, this MONTHLY, 74 (1967) 571–574.
5. Vera Pless, Introduction to the Theory of Error-Correcting Codes, John Wiley, New York, 1982.

# Rings With Invertible Regular Elements

F. C. LEARY
*Department of Mathematics, St. Bonaventure University, St. Bonaventure, NY 14778*

Let $R$ be a ring (assumed to have 1 and associative multiplication). Beginners in abstract algebra often find it annoying that $R$ may have the very unintegerlike property of having *zero divisors*, i.e., nonzero elements $x$ and $y$ for which $xy = 0$ (e.g., in $\mathbf{Z}/6\mathbf{Z}$, the ring of integers mod 6, $2 \cdot 3 = 0$). If $R$ is noncommutative, we call $x$ a *left zero divisor* and $y$ a *right zero divisor*. For commutative $R$, no such distinction is necessary.

Nonzero elements that are neither left nor right zero divisors are called *regular*. Regular elements resemble units (i.e., invertible elements) in that they may be cancelled. Of course, regular elements need not be units, e.g., the element 2 in $\mathbf{Z}$, the ring of integers. On the other hand, there are rings in which all regular elements are units, e.g., any division ring.

This note was motivated by class discussion relating to the following problem, which appears in [**3**, Chap. 20, Exer. C1]: Let $R$ be a finite commutative ring. Then every nonzero element of $R$ is either a zero divisor or a unit. The discussion centered around the fact that such an $R$ must then have the property $(P)$: all regular elements are units.

At this point, two questions naturally arise.

1. Do all finite rings have property $(P)$?
2. Is there a way to distinguish the rings that have property $(P)$ from those that do not?

The answer to (1) is yes and the proof is revealing.

First let $R$ be commutative and $x$ in $R$ be regular. Then left multiplication by $x$ provides an injective function $f: R \to R$. Since $R$ is a finite set, $f$ must also be surjective. Thus, $f(a) = xa = 1$ for some $a$ in $R$ and $x$ is a unit. This same argument applied to both left and right multiplication by $x$ shows that any finite ring has property $(P)$.

The key ingredient in the preceding discussion is the fact that for finite sets $S$, every injective $f: S \to S$ is necessarily surjective. A similar requirement on module homomorphisms is closely related to our discussion (see Theorem 1) and allows a characterization of commutative rings having property $(P)$.

Let $R$ be commutative and consider $R$ as an $R$-module. Any $R$-module homomorphism $f: R \to R$ is multiplication by an element $x$ of $R$. It is immediate that

multiplication by $x$ is injective if and only if $x$ is regular and surjective if and only if $x$ is a unit. These observations establish the following theorem.

THEOREM 1. *Let $R$ be commutative. Every regular element of $R$ is a unit if and only if every injective $R$-module homomorphism $f\colon R \to R$ is surjective.*

For arbitrary $R$, say that $R$ satisfies the *finiteness condition* if every injective (left/right) $R$-module homomorphism $f\colon R \to R$ is surjective. The finiteness condition is sufficient for $R$ to have property $(P)$ and unifies our previous observations: finite rings and division rings clearly satisfy the finiteness condition and so have property $(P)$; the commutative ring $\mathbf{Z}$ does not satisfy the condition (consider multiplication by 2) and so fails to have property $(P)$. In fact, an integral domain that is not a field cannot satisfy the finiteness condition and so cannot have property $(P)$. This is consistent with the result for finite rings because finite integral domains are fields. The following examples show that some very familiar rings from linear algebra have property $(P)$.

*Example 1.* Let $A$ be a finite dimensional algebra over the field $F$. An injective $F$-module homomorphism $T\colon A \to A$ must preserve independent sets and so take bases to bases. Therefore, $T$ is surjective and $A$ satisfies the finiteness condition. Hence, $A$ has property $(P)$.

As a special case, $M(n; F)$, the ring of $n \times n$ matrices with entries in $F$, satisfies the finiteness property and so has property $(P)$. Thus, the finiteness condition implies the familiar result that a homogeneous system of linear equations has only the trivial solution if and only if the matrix of coefficients is invertible. □

For noncommutative $R$, the finiteness condition actually implies more than property $(P)$. It implies that all right regular elements (i.e., nonzero elements that are not right zero divisors) are left invertible and that all left regular elements are right invertible. As it turns out, the finiteness condition is not necessary for $R$ to have $(P)$. To see this consider the von Neumann regular rings, those rings (not necessarily with 1) such that given $a$ in $R$ there is an $x$ in $R$ such that $axa = a$. Necessarily, von Neumann regular rings have property $(P)$ (consider $a$ regular and the two possibilities for factoring $axa - a = 0$). However, von Neumann regular rings need not satisfy the finiteness condition.

*Example 2.* Let $V$ be a vector space of countable dimension over a field $F$ and $R$ its ring of linear transformations. It is a standard exercise to show that $R$ is von Neumann regular. Let $\{x(i)\}$, $i = 1, 2, \ldots,$ be a basis for $V$ and consider the linear transformation $T\colon V \to V$ defined by $T[x(i)] = x(i+1)$ for all $i$. Then $T$ is clearly injective, but $T$ is not surjective since $x(1)$ is not in the range of $T$. Therefore, $R$ has property $(P)$ but does not satisfy the finiteness condition. □

To characterize the noncommutative $R$ with property $(P)$ we use a modified finiteness condition that considers left and right multiplication by elements of $R$ simultaneously rather than separately. Let $R^{op}$ be the opposite ring of $R$. As additive groups, $R$ and $R^{op}$ are identical. The multiplication in $R^{op}$ is defined by $a \circ b = ba$, where $ba$ is the product in $R$. The product ring $R \times R^{op}$ is a left $(R \times R^{op})$-module using ring multiplication for the scalar multiplication, i.e., $(r, s) \cdot (a, b) = (ra, s \circ b) = (ra, bs)$. As usual, any left $(R \times R^{op})$-module homomorphism $f\colon R \times R^{op} \to R \times R^{op}$ is right multiplication by $(a, b)$ for some $(a, b)$

in $R \times R^{op}$. Let $\Delta$ represent the set of homomorphisms $f$ determined by the elements $(a, a)$ in $R \times R^{op}$. Clearly, any such $f$ is injective if and only if $a$ is regular, and surjective if and only if $a$ is invertible.

THEOREM 2. *Let R be a ring. Every regular element of R is a unit if and only if every injective f in $\Delta$ is surjective.*

Note that the modified condition reduces to the original one if $R$ is commutative.

Rings with property $(P)$ are interesting in their own right because these are precisely the rings that are their own left quotient rings (see [**1**, **2**]). This fact reaffirms our earlier observation on integral domains because integral domains that are not fields are properly contained in their fields of fractions. On the other hand, semisimple left Artinian rings, such as $M(n; F)$, are their own left quotient rings (see [**2**]).

The interested reader might want to attempt the following exercises.

1. Let $R$ be commutative with 1. $R$ has property $(P)$ if and only if $R$ is not $R$-module isomorphic to a proper principal ideal;
2. Let $R$ be the ring of Example 2 and let $I$ be the ideal of transformations of finite rank. Put $S = R/I$. Show that $a$ in $S$ is regular if and only if $a = t + I$ where $t$ has finite dimensional kernel or $t$ has finite dimensional cokernel (the cokernel of $t$: $V \to V$ is $V/(\text{image of } t)$; the dimension of the cokernel is often referred to as the codimension of the image of $t$).

I would like to thank the referee of this paper for several valuable comments and especially for suggesting the finiteness condition as a unifying theme for the presentation. I would also like to thank E. Enochs for giving me the initial encouragement to write up these ideas.

REFERENCES

1. Carl Faith, Algebra: Rings, Modules and Categories I, Springer-Verlag, Berlin, 1973.
2. Thomas W. Hungerford, Algebra, Holt, Rinehart and Winston, New York, 1974.
3. Charles C. Pinter, A Book of Abstract Algebra, McGraw-Hill, New York, 1982.

E 3358. *Proposed by William P. Wardlaw, U.S. Naval Academy, Annapolis, MD.*

Suppose $b$, $r$, $s$, and $t$ are nonnegative integers satisfying $r + s + t = rb > 0$. Show that

$$\sum_{k=0}^{\lfloor s/b \rfloor} (-1)^k \binom{r}{k} \binom{s + r - 1 - bk}{s - bk} = \sum_{k=0}^{\lfloor t/b \rfloor} (-1)^k \binom{r}{k} \binom{t + r - 1 - bk}{t - bk}.$$

E 3359. *Proposed by Daniel Ullman, George Washington University, Washington, DC.*

Does there exist a finite increasing sequence $1 < a_1 < a_2 < \cdots < a_n$ of positive integers such that $\sum_{i=1}^n 1/a_i = 1$ and such that $a_i$ and $a_{i+1}$ are relatively prime for $i = 1, 2, \ldots, n - 1$?

E 3360. *Proposed by Steven Ratering, Central College, Pella, IA.*

Suppose

$$x_1 = 3,$$
$$2^{x_2} = 3^3,$$
$$2^{2^{x_3}} = 3^{3^3},$$
$$2^{2^{2^{x_4}}} = 3^{3^{3^3}},$$

and so on. Does $\lim_{n \to \infty} x_n$ exist?

## SOLUTIONS OF ELEMENTARY PROBLEMS

### Triangular Numbers Which Are Products of Two Other Triangular Numbers

E 3138 [1986, 215]. *Proposed by W. R. Utz, University of Missouri, Columbia.*

Let $T(n)$ denote the $n$th triangular number, i.e., $T(n) = n(n + 1)/2$.
(a) Show that there exist infinite sequences $a_n$ and $b_n$ such that

$$T(2) \cdot T(a_n) = T(b_n).$$

(b) Find pairs of triangular numbers whose product is triangular such that each member of the pair is arbitrarily large.
(c*) What is the asymptotic density of "composite" triangular numbers among all triangular numbers?

*Combined solution of (a) and (b) based primarily on the solution by the late Anders Bager, Hjørring, Denmark.* We obtain assertions (a) and (b) by establishing the following theorem.

THEOREM. *If $K$ is a positive integer that is not a perfect square, then the equation*

$$K \cdot T(x) = T(y) \tag{1}$$

*has infinitely many solutions in positive integers $x$, $y$.*

E 3358. *Proposed by William P. Wardlaw, U.S. Naval Academy, Annapolis, MD.*

Suppose $b$, $r$, $s$, and $t$ are nonnegative integers satisfying $r + s + t = rb > 0$. Show that

$$\sum_{k=0}^{\lfloor s/b \rfloor} (-1)^k \binom{r}{k} \binom{s + r - 1 - bk}{s - bk} = \sum_{k=0}^{\lfloor t/b \rfloor} (-1)^k \binom{r}{k} \binom{t + r - 1 - bk}{t - bk}.$$

E 3359. *Proposed by Daniel Ullman, George Washington University, Washington, DC.*

Does there exist a finite increasing sequence $1 < a_1 < a_2 < \cdots < a_n$ of positive integers such that $\sum_{i=1}^{n} 1/a_i = 1$ and such that $a_i$ and $a_{i+1}$ are relatively prime for $i = 1, 2, \ldots, n - 1$?

E 3360. *Proposed by Steven Ratering, Central College, Pella, IA.*

Suppose

$$x_1 = 3,$$
$$2^{x_2} = 3^3,$$
$$2^{2^{x_3}} = 3^{3^3},$$
$$2^{2^{2^{x_4}}} = 3^{3^{3^3}},$$

and so on. Does $\lim_{n \to \infty} x_n$ exist?

## SOLUTIONS OF ELEMENTARY PROBLEMS

### Triangular Numbers Which Are Products of Two Other Triangular Numbers

E 3138 [1986, 215]. *Proposed by W. R. Utz, University of Missouri, Columbia.*

Let $T(n)$ denote the $n$th triangular number, i.e., $T(n) = n(n + 1)/2$.

(a) Show that there exist infinite sequences $a_n$ and $b_n$ such that

$$T(2) \cdot T(a_n) = T(b_n).$$

(b) Find pairs of triangular numbers whose product is triangular such that each member of the pair is arbitrarily large.

(c*) What is the asymptotic density of "composite" triangular numbers among all triangular numbers?

*Combined solution of (a) and (b) based primarily on the solution by the late Anders Bager, Hjørring, Denmark.* We obtain assertions (a) and (b) by establishing the following theorem.

THEOREM. *If $K$ is a positive integer that is not a perfect square, then the equation*

$$K \cdot T(x) = T(y) \tag{1}$$

*has infinitely many solutions in positive integers $x$, $y$.*

*Remarks.* If $K = k^2$, where $k > 1$, the equation (1) is equivalent to

$$k^2(2x+1)^2 - (2y+1)^2 = k^2 - 1,$$

so that $k(2x+1) + (2y+1)$ and $k(2x+1) - (2y+1)$ must be complementary divisors of $k^2 - 1$. Thus when $K = k^2$, where $k > 1$, equation (1) has at most $d(k^2 - 1)$ solutions, where $d(j)$ is the number of divisors of the positive integer $j$.

As mentioned on page 33 of Volume 2 of Dickson's *History of the Theory of Numbers*, the result of the theorem goes back at least to A. Cunningham and R. W. D. Christie, *Math. Quest. Educ. Times*, 74 (1901) 87–88.

*Proof of Theorem.* Equation (1) is equivalent to

$$(2y+1)^2 - K(2x+1)^2 = 1 - K. \tag{2}$$

Let $(u_0, v_0)$ be the smallest positive integral solution of Pell's equation $u^2 - Kv^2 = 1$. If $u_0$ and $v_0$ are of opposite parity, we get infinitely many (but not necessarily all) positive integer solutions $x$, $y$ of (2) by taking

$$2y + 1 + (2x+1)\sqrt{K} = (1+\sqrt{K})(u_0 + v_0\sqrt{K})^j, \qquad j = 1, 2, 3, \dots;$$

that is, we express the right-hand side in the form $a_j + b_j\sqrt{K}$, where $a_j$ and $b_j$ are rational integers, and, observing that $a_j$ and $b_j$ are both odd, we put $2y + 1 = a_j$, $2x + 1 = b_j$.

If $u_0$ and $v_0$ are both odd (which can occur only when $K \equiv 0 \pmod 8$), we put

$$u_1 + v_1\sqrt{K} = (u_0 + v_0\sqrt{K})^2;$$

since $u_1$ is odd and $v_1$ is even, we get infinitely many positive integral solutions $x$, $y$ of (2) by taking

$$2y + 1 + (2x+1)\sqrt{K} = (1+\sqrt{K})(u_1 + v_1\sqrt{K})^j, \qquad j = 1, 2, 3 \dots.$$

This completes the proof of the theorem.

Part (a) follows from the theorem by taking $K = T(2) = 3$. Part (b) follows by taking $K = T(r)$ with $r$ large and $r \equiv 2, 3, 4, 5 \pmod 8$; for then $T(r) \equiv 2, 3 \pmod 4$ and hence $T(r)$ is not a square. More generally, $T(r)$ is not a square as long as $r$ is not in the sequence $c_1 = 1, c_2 = 8, c_3 = 49, c_4 = 288, \dots, c_{j+2} = 6c_{j+1} - c_j + 2$; for an argument with Pell's equation shows that $T(r)$ is a square if and only if

$$2r + 1 = \frac{1}{2}(3+\sqrt{8})^j + \frac{1}{2}(3-\sqrt{8})^j = 2c_j + 1$$

for some positive integer $j$. Thus we can obtain part (b) by taking $K = T(r)$, where $r$ is any large positive integer *not* in the rather sparse sequence $\{c_n\}$.

*Solution of (c) by Fred Dodd and Leon Mattics, University of S. Alabama, Mobile.* Presumably a triangular number $T(c)$ is regarded as "composite" if there exist positive integers $a$, $b$ greater than 1 such that $T(c) = T(a)T(b)$. The simplest examples of "composite" triangular numbers are $T(8) = T(3)T(3)$, $T(9) = T(2)T(5)$, $T(20) = T(4)T(6)$, $T(35) = T(2)T(20) = T(3)T(14)$, $T(39) = T(4)T(12)$, $T(44) = T(5)T(11)$.

The asymptotic density of the set of "composite" triangular numbers is zero. More specifically if $F(n)$ denotes the number of triples $a, b, c$ such that

$$T(a)T(b) = T(c), \qquad 1 < a \leqslant b < c \leqslant n, \tag{1}$$

we shall show that

$$F(n) < 4n^{3/4}. \tag{2}$$

We require the following lemma about hyperbolas.

LEMMA. *Suppose $g(x) = A\sqrt{x^2 - d^2}$ for $x \geqslant d$, where $A$ and $d$ are given positive numbers. Suppose $h$ is a fixed positive number. Then*

$$m_h(x) = \{g(x+h) + g(x-h)\}/(2x)$$

*is an increasing function of $x$ for $x \geqslant d + h$.*

*Remark.* Geometrically $m_h(x)$ is the slope of the line from the origin to the midpoint of the chord joining the points $(x - h, g(x - h))$ and $(x + h, g(x + h))$ on the hyperbola $y = g(x)$.

*Proof of the lemma.* Clearly

$$m_h(x) = \frac{g(x+h)^2 - g(x-h)^2}{2x\{g(x+h) - g(x-h)\}} = \frac{2A^2h}{g(x+h) - g(x-h)}.$$

Thus it suffices to show that $g(x + h) - g(x - h)$ is a decreasing function of $x$ for $x \geqslant d + h$. But for $x > d + h$ we have

$$g'(x+h) - g'(x-h) = \frac{A(x+h)}{\sqrt{(x+h)^2 - d^2}} - \frac{A(x-h)}{\sqrt{(x-h)^2 - d^2}} < 0,$$

since the derivative of $x(x^2 - d^2)^{-1/2}$ is $-d^2(x^2 - d^2)^{-3/2}$. Thus the lemma is proved.

Since $F(n) = 0$ for $n \leqslant 7$ we may assume $n \geqslant 8$. For given $a$ with $1 < a < n$ let $s(a, n)$ denote the number of pairs $b, c$ satisfying (1). If $b \geqslant a > 2^{1/4}n^{1/2}$, clearly $T(b) \geqslant T(a) > 2^{1/4}n^{1/2}(2^{1/4}n^{1/2} + 1)/2$ and, hence, $T(a)T(b) > n(n + 2^{3/4})/2$. Thus $s(a, n) = 0$ if $a > 2^{1/4}n^{1/2}$, and so

$$F(n) = \sum_{a=2}^{\lfloor 2^{1/4}n^{1/2} \rfloor} s(a, n).$$

Suppose that $a$ and $n$ are fixed and $s(a, n) > 0$. Set $K = T(a)$. Then the equation $T(a)T(b) = T(c)$ is equivalent to $K(b^2 + b) = c^2 + c$ or

$$K\left\{(2b+1)^2 - 1\right\} = (2c+1)^2 - 1.$$

Set $u = 2b + 1$, $v = 2c + 1$. Since $v^2 = (2c + 1)^2 \leqslant (2n + 1)^2$, we have

$$u^2 = \frac{v^2 - 1}{K} + 1 \leqslant \frac{4n^2 + 4n}{K} + 1 = 8\frac{n(n+1)}{a(a+1)} + 1 < 8\frac{n^2}{a^2} + 1 < 9\frac{n^2}{a^2},$$

so that

$$u < 3n/a. \tag{3}$$

On the other hand, $u = 2b + 1 \geqslant 2a + 1 = \sqrt{8K+1} > 2\sqrt{2K}$ and

$$v^2 = Ku^2 - K + 1 > K\{(2b+1)^2 - 1\} \geqslant K\{(2a+1)^2 - 1\} = 8K^2,$$

so that

$$0 < \sqrt{K}u - v = \frac{K-1}{\sqrt{K}u + v} < \frac{K}{2\sqrt{2}K + 2\sqrt{2}K}$$

or

$$0 < \sqrt{K}u - v < 1/(4\sqrt{2}). \tag{4}$$

Now suppose $(b_i, c_i)$, $i = 1, 2, \ldots, s$ are the solutions of $KT(b) = T(c)$ with $a \leqslant b_i < c_i \leqslant n$ and $b_1 < b_2 < \cdots < b_s$, where $s = s(a, n)$. Set $u_i = 2b_i + 1$ and $v_i = 2c_i + 1$ for $i = 1, 2, \ldots, s$. We claim that $u_{i+1} - u_i \neq u_{j+1} - u_j$ for $1 \leqslant i < j \leqslant s - 1$.

Suppose to the contrary that $u_{i+1} - u_i = u_{j+1} - u_j$ for some pair $i$, $j$ with $1 \leqslant i < j \leqslant s - 1$. From (4) we have

$$-1/(4\sqrt{2}) < \left(\sqrt{K}u_{i+1} - v_{i+1}\right) - \left(\sqrt{K}u_i - v_i\right) < 1/(4\sqrt{2}),$$

so that

$$\sqrt{K}(u_{i+1} - u_i) - (v_{i+1} - v_i) = \theta_i,$$

where $|\theta_i| < 1/(4\sqrt{2})$. Similarly

$$\sqrt{K}(u_{j+1} - u_j) - (v_{j+1} - v_j) = \theta_j,$$

where $|\theta_j| < 1/(4\sqrt{2})$. Hence,

$$v_{i+1} - v_i + \theta_i = \sqrt{K}(u_{i+1} - u_i) = \sqrt{K}(u_{j+1} - u_j) = v_{j+1} - v_j + \theta_j,$$

so that

$$|(v_{j+1} - v_j) - (v_{i+1} - v_i)| = |\theta_i - \theta_j| < 1/(2\sqrt{2}).$$

Since the left-hand side here is an integer, we have

$$v_{i+1} - v_i = v_{j+1} - v_j. \tag{5}$$

On the other hand $(u_i, v_i), (u_{i+1}, v_{i+1}), (u_j, v_j), (u_{j+1}, v_{j+1})$ are points with positive integer coordinates lying on the hyperbola $y^2 = Kx^2 - (K-1)$ and satisfying the conditions $u_{i+1} - u_i = u_{j+1} - u_j > 0$, $u_i < u_j$. Further,

$$\frac{v_{i+1} - v_i}{v_{j+1} - v_j} = \frac{K\left(u_{i+1}^2 - u_i^2\right)/(v_{i+1} + v_i)}{K\left(u_{j+1}^2 - u_j^2\right)/(v_{j+1} + v_j)} = \frac{(v_{j+1} + v_j)/(u_{j+1} + u_j)}{(v_{i+1} + v_i)/(u_{i+1} + u_i)}. \tag{6}$$

Applying the Lemma with $2h = u_{i+1} - u_i = u_{j+1} - u_j$ and $g(x) = \sqrt{Kx^2 - (K-1)}$, we find that

$$(v_{j+1} + v_j)/(u_{j+1} + u_j) > (v_{i+1} + v_i)/(u_{i+1} + u_i),$$

and then (6) gives $v_{i+1} - v_i > v_{j+1} - v_j$. But this contradicts (5) and so our assumption that $u_{i+1} - u_i = u_{j+1} - u_j$ is untenable.

Thus we have shown that the gaps $u_2 - u_1, u_3 - u_2, \ldots, u_s - u_{s-1}$ are $s - 1$ different even positive integers. Hence,

$$\begin{aligned} u_s - u_1 &= (u_2 - u_1) + (u_3 - u_2) + \cdots + (u_s - u_{s-1}) \\ &\geqslant 2 + 4 + \cdots + 2(s-1) = s(s-1). \end{aligned}$$

Combining this with (3), we obtain

$$3n/a > u_s > u_s - u_1 \geqslant (s-1)^2,$$

so that

$$s(a,n) < 1 + \sqrt{3n/a}.$$

Hence

$$\begin{aligned} F(n) &= \sum_{a=2}^{\lfloor 2^{1/4}n^{1/2}\rfloor} s(a,n) < \sum_{a=2}^{\lfloor 2^{1/4}n^{1/2}\rfloor} \left(1 + \sqrt{3n/a}\right) \\ &< 2^{1/4}n^{1/2} + \sqrt{3n}\int_1^{2^{1/4}n^{1/2}} t^{-1/2}\,dt \\ &< \sqrt{3n}\int_0^{2^{1/4}n^{1/2}} t^{-1/2}\,dt < 4n^{3/4}. \end{aligned}$$

Thus (2) is proved.

*Editorial comment.* The equation of Part (a) is equivalent to $(2b_n + 1)^2 - 3(2a_n + 1)^2 = -2$, which has the explicit solution

$$2b_n + 1 + (2a_n + 1)\sqrt{3} = (1 + \sqrt{3})(2 + \sqrt{3})^n, \qquad n = 1, 2, 3, \ldots$$

or the recursive solution $a_1 = 1$, $a_2 = 5$, $a_{n+2} = 4a_{n+1} - a_n + 1$, $b_1 = 2$, $b_2 = 9$, $b_{n+2} = 4b_{n+1} - b_n + 1$. The solution can also be given in terms of the continued fraction expansion of $\sqrt{3}$.

Solutions of (b) along the same lines as Bager's were given by László Cseh, Emil Grosswald, Roger B. Nelsen, Donald Redmond, and Paul Tzermias.

Fred Dodd and Leon Mattics, Walther Janous, and Michael Vowe gave solutions of (b) by noting that $T(x)T(y) = T(x^2 + x)$ is equivalent to $T(y) = 2T(x) + 1$, which reduces to a Pell equation and has infinitely many solutions given by

$$2y + 1 + (2x + 1)\sqrt{2} = (3 \pm \sqrt{2})(3 + 2\sqrt{2})^j, \qquad j = 1, 2, 3, \ldots.$$

As mentioned on page 35 of Volume 2 of Dickson's *History of the Theory of Numbers*, this argument goes back at least to A. Gérardin, Sphinx-Oedipe 8 (1913) 128. Similarly H. Y. Lin and Byron Siu (jointly) and also Daniel Ullman gave solutions of (b) by noting that $T(x)T(y) = T(x^2 - 1)$ is equivalent to $T(y) = 2T(x - 1)$, and John P. Robertson gave a solution of (b) by noting that $T(x)T(y) = T(T(y))$ is equivalent to $T(y) = 2T(x) - 1$.

Parts (a) and (b) were solved by A. Bagers, R. Bernstein, L. Cseh, F. Dodd and L. Mattics, E. Grosswald, W. Janous, O. P. Lossers, R. B. Nelsen, D. Redmond, J. P. Robertson, H. Y. Lin and B. Siu, P. Tzermias, D. Ullman, and M. Vowe. I. Paasche gave a solution of Part (a). Cseh remarked that there are infinitely many triangular numbers expressible in more than one way as a product of two triangular numbers greater than 1 (cf. *Elemente der Math.*, 26(1971) 68–69).

The above solution for part (c) was the only one received before the customary deadline. In response to an inquiry from the editors, Professor Andrzej Schinzel of the Polish Academy of Sciences produced a proof that the Dodd-Mattics inequality (2) can be replaced by the stronger estimate

$$F(n) = O\left(n^{1/2}(\log n)^2\right).$$

Professor Schinzel also gave a proof that $\log\log F(n) > (\log 2)(\log\log n)/(\log\log\log n)$ if $n$ is sufficiently large. It would be reasonable to conjecture that $F(n)$ actually grows faster than a certain positive power of $n$, i.e., that

$$\log\log F(n) > \log\log n - c$$

for sufficiently large $n$, where $c$ is a suitable positive constant; but this seems difficult to prove.

Then $m(H)/m(Q) > \delta/\varepsilon > \delta$, and, hence,

$$m(A)/m(H) < (1+\varepsilon)m(A)/m(\Omega) = \varepsilon$$

for every $A \in \Sigma$. This completes the proof of the lemma.

We now prove the main result claimed above. Put $Q = Q_0$ and suppose, for convenience, that $Q_0$ has area 1 (rather than area 4). Associate $H_0$ to $Q_0$ as in the lemma. The remainder, $Q_0 \setminus H_0$, has area $m(Q_0 \setminus H_0) < 1 - \delta$ and is a finite union of nonoverlapping rectangles $Q_i$. Put an $H_i$ into each $Q_i$ in exactly the same way as in the lemma. The area of the part of $Q_0$ that is not covered by any $H_i$ chosen so far is less than $(1-\delta)^2$, and the uncovered part is a finite union of nonoverlapping rectangles, into each of which we put an $H$ as in the lemma, and so on. After finitely many steps we obtain nonoverlapping sets $H_0, H_1, \dots, H_N$ (each of which is a finite collection of rectangles) whose union covers all of $Q_0$ except for a set $E$ with $m(E) < (1-\delta)^p < \varepsilon$, assuming $p$ is large enough.

We let $S$ be the collection of all rectangles $A$ that occur in any $H_i$, and in addition those (of total area less than $\varepsilon$) that cover $E$. Any nonoverlapping subcollection $\tilde{S}$ of $S$ contains at most one $A$ from any $H$. By (iii) of the lemma, the area of the union of those $A$'s is less than $\varepsilon$. Adding the contribution of $E$, we conclude that the union of the members of $\tilde{S}$ has area less than $2\varepsilon$.

The theorem holds a fortiori in $R^n$, for any $n \geqslant 2$.

*Editorial comment.* Daniel Ullman pointed out that the argument of Solution I also extends to higher dimensions, but that the answer to the existence question is always "yes" in $\mathbb{R}^1$. A referee mentions R. Rado, *Proc. London Math. Soc.*, (2) 51(1950) 232–264 and M. Ajtai, *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astron. Phys.*, 21(1973) 61–63 for similar problems concerning pairwise disjoint subfamilies of squares not constrained to cover a particular rectangle.

Solved also (with the generalization to higher dimensions) by J. Isbell, O. P. Lossers (The Netherlands), D. Ullman, and partially by the proposer.

### A Function Derived from Euler's Totient

E 3246 [1988, 51]. *Proposed by Solomon W. Golomb, University of Southern California, Los Angeles, CA.*

Define a function $\psi$ on the positive integers by putting $\psi(n) = n\phi(n)$, where $\phi$ is Euler's function. (a) Is $\psi$ an injective function? (b) For how many positive integers $n$ is $\psi(n)$ a perfect square? (c) Given any positive integer $m$, prove that there is an integer $n$ divisible by $m$ such that $\psi(n)$ is a perfect cube.

*Solution by an anonymous contributor.* The answer to (a) is yes, and the only positive integer $n$ for which $\psi(n)$ is a perfect square is $n = 1$. We prove the following more general result:

THEOREM. *Let $\psi_k(n) = n^k\phi(n)$, where $k$ is a fixed positive integer and $\phi$ is Euler's function. Then*

(a) *$\psi_k$ is injective.*

(b) *If $s$ is an integer such that* $\gcd(k+1, s) > 1$, *then $\psi_k(n)$ is the $s$th power of an integer only for $n = 1$.*

Then $m(H)/m(Q) > \delta/\varepsilon > \delta$, and, hence,

$$m(A)/m(H) < (1+\varepsilon)m(A)/m(\Omega) = \varepsilon$$

for every $A \in \Sigma$. This completes the proof of the lemma.

We now prove the main result claimed above. Put $Q = Q_0$ and suppose, for convenience, that $Q_0$ has area 1 (rather than area 4). Associate $H_0$ to $Q_0$ as in the lemma. The remainder, $Q_0 \setminus H_0$, has area $m(Q_0 \setminus H_0) < 1 - \delta$ and is a finite union of nonoverlapping rectangles $Q_i$. Put an $H_i$ into each $Q_i$ in exactly the same way as in the lemma. The area of the part of $Q_0$ that is not covered by any $H_i$ chosen so far is less than $(1-\delta)^2$, and the uncovered part is a finite union of nonoverlapping rectangles, into each of which we put an $H$ as in the lemma, and so on. After finitely many steps we obtain nonoverlapping sets $H_0, H_1, \dots, H_N$ (each of which is a finite collection of rectangles) whose union covers all of $Q_0$ except for a set $E$ with $m(E) < (1-\delta)^p < \varepsilon$, assuming $p$ is large enough.

We let $S$ be the collection of all rectangles $A$ that occur in any $H_i$, and in addition those (of total area less than $\varepsilon$) that cover $E$. Any nonoverlapping subcollection $\tilde{S}$ of $S$ contains at most one $A$ from any $H$. By (iii) of the lemma, the area of the union of those $A$'s is less than $\varepsilon$. Adding the contribution of $E$, we conclude that the union of the members of $\tilde{S}$ has area less than $2\varepsilon$.

The theorem holds a fortiori in $R^n$, for any $n \geqslant 2$.

*Editorial comment.* Daniel Ullman pointed out that the argument of Solution I also extends to higher dimensions, but that the answer to the existence question is always "yes" in $\mathbb{R}^1$. A referee mentions R. Rado, *Proc. London Math. Soc.*, (2) 51(1950) 232–264 and M. Ajtai, *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astron. Phys.*, 21(1973) 61–63 for similar problems concerning pairwise disjoint subfamilies of squares not constrained to cover a particular rectangle.

Solved also (with the generalization to higher dimensions) by J. Isbell, O. P. Lossers (The Netherlands), D. Ullman, and partially by the proposer.

### A Function Derived from Euler's Totient

E 3246 [1988, 51]. *Proposed by Solomon W. Golomb, University of Southern California, Los Angeles, CA.*

Define a function $\psi$ on the positive integers by putting $\psi(n) = n\phi(n)$, where $\phi$ is Euler's function. (a) Is $\psi$ an injective function? (b) For how many positive integers $n$ is $\psi(n)$ a perfect square? (c) Given any positive integer $m$, prove that there is an integer $n$ divisible by $m$ such that $\psi(n)$ is a perfect cube.

*Solution by an anonymous contributor.* The answer to (a) is yes, and the only positive integer $n$ for which $\psi(n)$ is a perfect square is $n = 1$. We prove the following more general result:

THEOREM. *Let $\psi_k(n) = n^k\phi(n)$, where $k$ is a fixed positive integer and $\phi$ is Euler's function. Then*

(a) *$\psi_k$ is injective.*

(b) *If $s$ is an integer such that* $\gcd(k+1, s) > 1$, *then $\psi_k(n)$ is the $s$th power of an integer only for $n = 1$.*

## POSTMORTEM COMMENTS ON ELEMENTARY PROBLEMS

### Distances with Specified Multiplicities

E 2938* [1982, 273; 1989, 447]. *Proposed by P. Erdös, Hungarian Academy of Sciences.*

Can one find $n$ points in the plane (no three on a line, no four on a circle) so that for every $i$, $i = 1, 2, \ldots, n-1$, there is a distance determined by these points that occurs exactly $i$ times?

*Correction to previous editorial comment.* Roman W. Wong has pointed out that the example given for $n = 4$ in the May issue does not meet the condition that no three points lie on a line and suggests replacing the sentence concerning $n = 4$ by the following:

"For $n = 4$ consider the points $(-1, 0)$, $(0, -1)$, $(1, 0)$, and $(0, \sqrt{3})$; the distances $1 + \sqrt{3}$, $\sqrt{2}$, and 2 occur once, twice, and three times respectively."

A similar remark was made by Barbara Kreaseck.

For large values of $n$ the problem is still open.

### A Minimum Under a Constraint

E 3099 [1985, 507; 1987, 1003]. *Proposed by Weixuan Li and Edward T. H. Wang, Wilfrid Laurier University, Waterloo, Ontario, Canada.*

Let $a_1 \leqslant a_2 \leqslant \cdots \leqslant a_n$ be $n$ nonnegative reals ($n \geqslant 2$) such that

$$\sum_{i=1}^{n} a_i a_{i+1} = 1 \qquad (a_{n+1} = a_1).$$

Determine the minimum value of $\sum_{i=1}^{n} a_i$.

*Editorial comment.* Professor Mark Ashbaugh observed that the published solution was less than complete, because it used Lagrange multipliers without qualification in the presence of boundary extrema. Further, he noted the troublesome boundary conditions $a_1 \leqslant a_2 \leqslant \cdots \leqslant a_n$ are actually irrelevant for the value of the minimum. The following solution uses identities in place of Lagrange multipliers.

*Composite solution by Mark S. Ashbaugh and by Zun Shan and Edward T. H. Wang.* We shall prove the following theorem.

THEOREM. *If $n$ is a positive integer greater than* 1, *let*

$$M_n = \min \sum_{i=1}^{n} a_i,$$

*where the minimum is taken over all $n$-tuples of nonnegative real numbers $a_1, a_2, \ldots, a_n$ satisfying*

$$\sum_{i=1}^{n} a_i a_{i+1} = 1 \qquad (a_{n+1} = a_1).$$

*Then $M_2 = \sqrt{2}$, $M_3 = \sqrt{3}$, and $M_n = 2$ for all $n \geqslant 4$.*

## POSTMORTEM COMMENTS ON ELEMENTARY PROBLEMS

### Distances with Specified Multiplicities

E 2938* [1982, 273; 1989, 447]. *Proposed by P. Erdös, Hungarian Academy of Sciences.*

Can one find $n$ points in the plane (no three on a line, no four on a circle) so that for every $i$, $i = 1, 2, \ldots, n-1$, there is a distance determined by these points that occurs exactly $i$ times?

*Correction to previous editorial comment.* Roman W. Wong has pointed out that the example given for $n = 4$ in the May issue does not meet the condition that no three points lie on a line and suggests replacing the sentence concerning $n = 4$ by the following:

"For $n = 4$ consider the points $(-1, 0)$, $(0, -1)$, $(1, 0)$, and $(0, \sqrt{3})$; the distances $1 + \sqrt{3}$, $\sqrt{2}$, and 2 occur once, twice, and three times respectively."

A similar remark was made by Barbara Kreaseck.

For large values of $n$ the problem is still open.

### A Minimum Under a Constraint

E 3099 [1985, 507; 1987, 1003]. *Proposed by Weixuan Li and Edward T. H. Wang, Wilfrid Laurier University, Waterloo, Ontario, Canada.*

Let $a_1 \leqslant a_2 \leqslant \cdots \leqslant a_n$ be $n$ nonnegative reals ($n \geqslant 2$) such that

$$\sum_{i=1}^{n} a_i a_{i+1} = 1 \qquad (a_{n+1} = a_1).$$

Determine the minimum value of $\sum_{i=1}^{n} a_i$.

*Editorial comment.* Professor Mark Ashbaugh observed that the published solution was less than complete, because it used Lagrange multipliers without qualification in the presence of boundary extrema. Further, he noted the troublesome boundary conditions $a_1 \leqslant a_2 \leqslant \cdots \leqslant a_n$ are actually irrelevant for the value of the minimum. The following solution uses identities in place of Lagrange multipliers.

*Composite solution by Mark S. Ashbaugh and by Zun Shan and Edward T. H. Wang.* We shall prove the following theorem.

THEOREM. *If $n$ is a positive integer greater than* 1, *let*

$$M_n = \min \sum_{i=1}^{n} a_i,$$

*where the minimum is taken over all $n$-tuples of nonnegative real numbers $a_1, a_2, \ldots, a_n$ satisfying*

$$\sum_{i=1}^{n} a_i a_{i+1} = 1 \qquad (a_{n+1} = a_1).$$

*Then $M_2 = \sqrt{2}$, $M_3 = \sqrt{3}$, and $M_n = 2$ for all $n \geqslant 4$.*

### Trapezoids in a Polygon

E 3157 [1986, 482; 1988, 557]. *Proposed by Liviu I. Nicoaescu, Iassy, Romania.*

How many sets of four distinct points forming the vertices of a trapezoid are there if the points are chosen from the vertices of a regular $n$-gon ($n \geqslant 4$)?

*Editorial comment.* The published solution (and the solutions of the nine other persons listed as solvers) interpreted the term trapezoid in the usual way, i.e., to mean "a convex quadrilateral with at least one pair of parallel sides" and obtained the count $n(n-1)(n-3)/8$ for $n$ odd and $n(n-2)(n-3)/8$ for $n$ even.

Two other readers, Robert E. Bernstein and Allan Pedersen, interpreted the term trapezoid to mean "a convex quadrilateral with exactly one pair of parallel sides" and thus obtained a different result for $n$ even; however, their results are correct for this altered interpretation.

### Concurrent Conics

E 3172 [1986, 733; 1988, 957]. *Proposed by Jordi Dou, Barcelona, Spain.*

Let $A'$ (respectively $B'$, $C'$) be the foot of the altitude from vertex $A$ (respectively $B$, $C$) in a triangle $ABC$. Let $H$ be its orthocenter, and let $M$ be an arbitrary point of the plane. Prove that the conics $MABA'B'$, $MBCB'C'$, $MCAC'A'$, $MHCA'B'$, $MHAB'C'$, and $MHBC'A'$ have a point other than $M$ in common.

*Remark by Peter Yff, Ball State University, Muncie, IN.* Part of the hypothesis of the problem is redundant. Since all properties of the figure are preserved under projection, $H$ need not be the orthocenter of $ABC$. In fact $A$, $B$, $C$, $H$ can be any four points no three of which are collinear.

This is implicit in the published solution by Lossers, where "projective coordinates" are mentioned. In fact, Lossers' entire proof holds verbatim for any position of $H$. Only his remark that $M$ and $N$ are isogonal conjugates with respect to $A'B'C'$ (which is not called for in the statement of the problem) needs to be altered. The coordinates of $N$ are still $(\mu_1^{-1}, \mu_2^{-1}, \mu_3^{-1})$, but this makes $N$ the isogonal conjugate of $M$ only if the coordinates are trilinear, i.e., if $H = (1,1,1)$ is the incenter of $A'B'C'$ (the orthocenter of $ABC$). If $ABC$ is obtuse, its orthocenter is an excenter of $A'B'C'$, but this does not affect isogonal conjugacy.

### Maximal Sets of Dispersed Vectors

E 3206 [1987, 373; 1989, 161]. *Proposed by Chico Problem Group, California State University, Chico.*

Let $S = \{x_1, \dots, x_n\}$ be a set of $n$ nonzero vectors in $\mathbb{R}^d$ such that $x_i x_j \leqslant 0$ whenever $1 \leqslant i < j \leqslant n$.

(a) Prove that $n \leqslant 2d$ and characterize those $S$ for which $n = 2d$.

(b) What if $x_i x_j \leqslant 0$ is replaced by $x_i x_j < 0$?

*Editorial comment.* If $x_1, \dots, x_n$ are vectors in $\mathbb{R}^d$ *of length one*, then the result of (a) takes the form: If $n \geqslant 2d + 1$, there exists $i$, $j$ with $1 \leqslant i < j \leqslant n$ such that

$\|x_i - x_j\|^2 < 2$; if $n = 2d$ and $\|x_i - x_j\|^2 \geqslant 2$ whenever $1 \leqslant i < j \leqslant n$, then $\{x_1, \ldots, x_n\}$ spans a cross-polytope ($n$-dimensional octahedron). In a letter to the editor, Professor László A. Székely remarked that in this form the assertion of (a) is frequently referred to as the Hajós-Davenport Lemma.

Professor Székely also mentioned that part (a) occurred as Problem 20 in *Matematikai Lapok* [2(1951)76] and that part (b) occurred as Problem 35 in *Matematikai Lapok* [2(1951)35; 3(1952)94]. Further relevant references are

D. G. Larman and C. A. Rogers, The realization of distances within sets in Euclidean space, *Mathematika*, 19(1972) 1–24, particularly p. 12.

R. A. Rankin, The closest packing of spherical caps in $n$-dimensions, *Proc. Glasgow Math. Assoc.* 2(1955) 139–144.

### How to Gamble if You Must

E 3219 [1987, 680; 1989, 163]. *Proposed by Daniel Rawsthorne, Silver Spring, MD.*

A gambler has no money, but the host of the casino generously allows him to play 100 games of the following type. He may either (1) choose to accept one dollar with no risk or (2) choose an integer $n > 1$, whereupon he wins $n$ dollars with probability $2/(n+1)$ or loses one dollar with probability $(n-1)/(n+1)$. (He must have at least one dollar to choose option (2).) What is an optimal strategy for the gambler if he wishes to leave with \$200 or more, and what is his probability of success using that strategy?

*Editorial comment.* It has come to the attention of the editors that the solution published in the February, 1989 issue is partially incorrect. The strategy given there is optimal in the sense that it has the maximum possible success probability (namely, $1/2$), but the accompanying assertion that it is the only such strategy is very far from the truth.

The strategy used by most solvers, including the author of the published solution, may be schematically indicated by the 100-tuple

$$(1, 101, 1, 103, 1, 105, \ldots, 1, 195, 1, 197, 1, 199).$$

Here the ones in the odd-numbered places indicate a use of option (1) in the odd-numbered games; the numbers $101, 103, \ldots, 199$ indicate the successive use of option (2) in the even-numbered games with $n = 101, 103, \ldots, 199$ respectively, with the understanding that in the event of a win using option (2) the gambler reverts to option (1) for all remaining games. However, an equally effective strategy would be that similarly indicated by any 100-tuple involving fifty ones and the fifty numbers $101, 103, \ldots, 199$ which satisfies the following two conditions:

(i) 101 occurs before 103, 103 occurs before 105, 105 occurs before 107, etc.,
(ii) for $1 \leqslant k \leqslant 100$ the number of ones occurring in the first $k$ places of the 100-tuple must be at least as great as the number of integers taken from the set $\{101, 103, \ldots, 199\}$ occurring in the first $k$ places.

$\|x_i - x_j\|^2 < 2$; if $n = 2d$ and $\|x_i - x_j\|^2 \geqslant 2$ whenever $1 \leqslant i < j \leqslant n$, then $\{x_1, \ldots, x_n\}$ spans a cross-polytope ($n$-dimensional octahedron). In a letter to the editor, Professor László A. Székely remarked that in this form the assertion of (a) is frequently referred to as the Hajós-Davenport Lemma.

Professor Székely also mentioned that part (a) occurred as Problem 20 in *Matematikai Lapok* [2(1951)76] and that part (b) occurred as Problem 35 in *Matematikai Lapok* [2(1951)35; 3(1952)94]. Further relevant references are

D. G. Larman and C. A. Rogers, The realization of distances within sets in Euclidean space, *Mathematika*, 19(1972) 1–24, particularly p. 12.

R. A. Rankin, The closest packing of spherical caps in $n$-dimensions, *Proc. Glasgow Math. Assoc.* 2(1955) 139–144.

### How to Gamble if You Must

E 3219 [1987, 680; 1989, 163]. *Proposed by Daniel Rawsthorne, Silver Spring, MD.*

A gambler has no money, but the host of the casino generously allows him to play 100 games of the following type. He may either (1) choose to accept one dollar with no risk or (2) choose an integer $n > 1$, whereupon he wins $n$ dollars with probability $2/(n+1)$ or loses one dollar with probability $(n-1)/(n+1)$. (He must have at least one dollar to choose option (2).) What is an optimal strategy for the gambler if he wishes to leave with \$200 or more, and what is his probability of success using that strategy?

*Editorial comment.* It has come to the attention of the editors that the solution published in the February, 1989 issue is partially incorrect. The strategy given there is optimal in the sense that it has the maximum possible success probability (namely, $1/2$), but the accompanying assertion that it is the only such strategy is very far from the truth.

The strategy used by most solvers, including the author of the published solution, may be schematically indicated by the 100-tuple

$$(1, 101, 1, 103, 1, 105, \ldots, 1, 195, 1, 197, 1, 199).$$

Here the ones in the odd-numbered places indicate a use of option (1) in the odd-numbered games; the numbers $101, 103, \ldots, 199$ indicate the successive use of option (2) in the even-numbered games with $n = 101, 103, \ldots, 199$ respectively, with the understanding that in the event of a win using option (2) the gambler reverts to option (1) for all remaining games. However, an equally effective strategy would be that similarly indicated by any 100-tuple involving fifty ones and the fifty numbers $101, 103, \ldots, 199$ which satisfies the following two conditions:

(i) 101 occurs before 103, 103 occurs before 105, 105 occurs before 107, etc.,

(ii) for $1 \leqslant k \leqslant 100$ the number of ones occurring in the first $k$ places of the 100-tuple must be at least as great as the number of integers taken from the set $\{101, 103, \ldots, 199\}$ occurring in the first $k$ places.

## ADVANCED PROBLEMS

6616. *Proposed by Hugh M. W. Edgar, San Jose State University, CA.*

Let $d(n)$ denote the number of positive integral divisors of $n$ and let $\sigma(n)$ denote the sum of these divisors. Let $S$ be the set of positive integers with exactly two distinct prime factors (repeated prime factors are permitted). For $n \in S$ prove that the following three assertions are equivalent:

(1) $n$ is an even perfect number, i.e., $n$ is even and $\sigma(n) = 2n$.

(2) the harmonic mean of the divisors of $n$ is integral, i.e., $nd(n)/\sigma(n)$ is an integer.

(3) $\sigma(n)$ has exactly the same prime factors as $n$.

6617. *Proposed by A. Vince, University of Florida, Gainesville.*

A graph $\Gamma$ is *regular* if each vertex has the same degree. For a vertex $x$ let $\Gamma_x$ and $\Delta_x$ denote the subgraphs of $\Gamma - x$ induced by the vertices adjacent to and nonadjacent to $x$, respectively. Define *superregular* recursively as follows. The empty graph is superregular and $\Gamma$ is superregular if $\Gamma$ is regular and both $\Gamma_x$ and $\Delta_x$ are superregular for all $x$. Characterize the superregular graphs.

6618. *Proposed by Armel Mercier, Université du Québec à Chicoutimi, Canada.*

Let $\omega(n)$ denote the number of distinct prime factors of $n$, $p(n)$ the smallest prime factor of $n$, and $P_m(n)$ the $m$th largest prime factor of $n$.

(i) Prove that for any nonnegative integer $m$ we have

$$P_{m+1}(n) = (-1)^{m+1} \sum_{d\mid n} \binom{\omega(d) - 1}{m} \mu(d) p(d),$$

where $\mu$ denotes the Möbius function.

(ii) Prove that for any nonnegative integer $m$ we have

$$\sum_{d\mid n} \{\omega(d)\}^m \mu(d) p(d) = \sum_{j=0}^{m} (-1)^{j+1} j! S(m+1, j+1) P_{j+1}(n),$$

where $S(m, j)$ denotes the Stirling number of the second kind.

## SOLUTIONS OF ADVANCED PROBLEMS

6563 [1988, 60]. *Proposed by C. A. Spiro, SUNY at Buffalo.*

Prove that, if $N$ is sufficiently large, at least one of the integers $N, N+1, \ldots, N+9$ has more than two distinct prime factors.

*Solution by J. Herzog, T. Maxsein, and P. R. Smith, Johann Wolfgang Goethe Universität, Frankfurt am Main, West Germany.* For a positive integer $N$ let $M_N$ be the set $\{N, N+1, \ldots, N+9\}$. We shall prove that if $N \geqslant 93$, then at least one element of $M_N$ has more than two distinct prime factors. For the proof we need the following elementary lemma on exponential diophantine equations.

## ADVANCED PROBLEMS

6616. *Proposed by Hugh M. W. Edgar, San Jose State University, CA.*

Let $d(n)$ denote the number of positive integral divisors of $n$ and let $\sigma(n)$ denote the sum of these divisors. Let $S$ be the set of positive integers with exactly two distinct prime factors (repeated prime factors are permitted). For $n \in S$ prove that the following three assertions are equivalent:

(1) $n$ is an even perfect number, i.e., $n$ is even and $\sigma(n) = 2n$.

(2) the harmonic mean of the divisors of $n$ is integral, i.e., $nd(n)/\sigma(n)$ is an integer.

(3) $\sigma(n)$ has exactly the same prime factors as $n$.

6617. *Proposed by A. Vince, University of Florida, Gainesville.*

A graph $\Gamma$ is *regular* if each vertex has the same degree. For a vertex $x$ let $\Gamma_x$ and $\Delta_x$ denote the subgraphs of $\Gamma - x$ induced by the vertices adjacent to and nonadjacent to $x$, respectively. Define *superregular* recursively as follows. The empty graph is superregular and $\Gamma$ is superregular if $\Gamma$ is regular and both $\Gamma_x$ and $\Delta_x$ are superregular for all $x$. Characterize the superregular graphs.

6618. *Proposed by Armel Mercier, Université du Québec à Chicoutimi, Canada.*

Let $\omega(n)$ denote the number of distinct prime factors of $n$, $p(n)$ the smallest prime factor of $n$, and $P_m(n)$ the $m$th largest prime factor of $n$.

(i) Prove that for any nonnegative integer $m$ we have

$$P_{m+1}(n) = (-1)^{m+1} \sum_{d \mid n} \binom{\omega(d) - 1}{m} \mu(d) p(d),$$

where $\mu$ denotes the Möbius function.

(ii) Prove that for any nonnegative integer $m$ we have

$$\sum_{d \mid n} \{\omega(d)\}^m \mu(d) p(d) = \sum_{j=0}^{m} (-1)^{j+1} j! S(m+1, j+1) P_{j+1}(n),$$

where $S(m, j)$ denotes the Stirling number of the second kind.

## SOLUTIONS OF ADVANCED PROBLEMS

6563 [1988, 60]. *Proposed by C. A. Spiro, SUNY at Buffalo.*

Prove that, if $N$ is sufficiently large, at least one of the integers $N, N+1, \ldots, N+9$ has more than two distinct prime factors.

*Solution by J. Herzog, T. Maxsein, and P. R. Smith, Johann Wolfgang Goethe Universität, Frankfurt am Main, West Germany.* For a positive integer $N$ let $M_N$ be the set $\{N, N+1, \ldots, N+9\}$. We shall prove that if $N \geqslant 93$, then at least one element of $M_N$ has more than two distinct prime factors. For the proof we need the following elementary lemma on exponential diophantine equations.

LEMMA. *Suppose* $\alpha, \beta, \gamma, \delta$ *are positive integers. Then*
(A) $2^\alpha 3^\beta - 4 = 2^\gamma 5^\delta$ iff $(\alpha,\beta,\gamma,\delta) = (1,3,1,2),(2,4,6,1),(3,1,2,1)$;
(B) $2^\alpha 3^\beta - 2 = 2^\gamma 5^\delta$ iff $(\alpha,\beta,\gamma,\delta) = (1,4,5,1),(2,1,1,1)$;
(C) $2^\alpha 3^\beta + 2 = 2^\gamma 5^\delta$ iff $(\alpha,\beta,\gamma,\delta) = (1,2,2,1),(4,1,1,2)$;
(D) $2^\alpha 3^\beta + 4 = 2^\gamma 5^\delta$ iff $(\alpha,\beta,\gamma,\delta) = (1,1,1,1),(2,2,3,1),(5,1,2,2)$;
(E) $2^\alpha + 1 = 3^\beta$ iff $(\alpha,\beta) = (1,1),(3,2)$;
(F) $2^\alpha - 1 = 3^\beta$ iff $(\alpha,\beta) = (2,1)$.

*Proof.* The results of (B) and (C) follow from those of (A) and (D) respectively by multiplying (B) and (C) by 2. Thus it suffices to prove (A), (D), (E), and (F). We give complete proofs for (F), (E), and (A), but leave the proof of (D) to the reader.

Suppose $\beta \geqslant 2$ in (F). Since 2 is a primitive root modulo 9 (and therefore modulo $3^\beta$) and since $\phi(3^\beta) = 2 \cdot 3^{\beta-1}$, it follows that $\alpha \equiv 0 (\bmod 2 \cdot 3^{\beta-1})$. Thus

$$2^\alpha \geqslant 2^{2\cdot 3^{\beta-1}} > 4 \cdot 3^{\beta-1} > 3^\beta + 1,$$

since $2^n > 2n$ for $n > 2$. Thus the only solution of (F) is that listed. (For a proof that, if $p$ is an odd prime, any primitive root modulo $p^2$ is automatically a primitive root modulo $p^r$, for any positive integer $r$, see [**3**, §17].)

Suppose $\beta \geqslant 3$ in (E). Since 2 is a primitive root modulo $3^\beta$, we have $2^{3^{\beta-1}} \equiv -1 (\bmod 3^\beta)$ and so $\alpha \equiv 3^{\beta-1} (\bmod 2 \cdot 3^{\beta-1})$. Then

$$2^\alpha + 1 > 2^\alpha \geqslant 2^{3^{\beta-1}} > 4 \cdot 3^{\beta-1} > 3^\beta,$$

since $2^n > 4n$ for $n > 4$. Thus the only solutions of (E) are those listed.

We now turn to the proof of (A). We distinguish three cases, namely, $\alpha = 1$, $\alpha = 2$, and $\alpha > 2$.

Case 1, $\alpha = 1$. Clearly $\gamma = 1$. Thus the equation of (A) becomes $3^\beta - 2 = 5^\delta$. Suppose $\beta > 3$. Since 5 is a primitive root modulo 81 and $5^{20} \equiv -2 (\bmod 81)$, we must have $\delta \equiv 20 (\bmod 54)$. Thus $5^\delta \equiv 5^{20} \equiv 5^2 (\bmod 19)$ and so $3^\beta = 2 + 5^\delta \equiv 3^3 (\bmod 19)$; since 3 is a primitive root modulo 19, we therefore have $\beta \equiv 3 (\bmod 18)$. We now consider $5^\delta$ and $3^\beta$ modulo 271. Since 5 has multiplicative order 27 modulo 271 and since $\delta \equiv 20 (\bmod 27)$, we have $5^\delta \equiv 5^{20} \equiv 88 (\bmod 271)$; thus $3^\beta = 2 + 5^\delta \equiv 90 (\bmod 271)$. But 3 has multiplicative order 30 modulo 271 and $90 \equiv 3^{14} (\bmod 271)$, so that $\beta \equiv 14 (\bmod 30)$. But the two congruences $\beta \equiv 3 (\bmod 18)$ and $\beta \equiv 14 (\bmod 30)$ are incompatible. Thus $\beta > 3$ is impossible and so the only solution of (A) in case 1 is $(1,3,1,2)$.

Case 2, $\alpha = 2$. Clearly $\gamma > 2$ and then the equation of (A) becomes $3^\beta - 1 = 2^{\gamma-2}5^\delta$. Accordingly $3^\beta \equiv 1 (\bmod 5)$ and so $\beta \equiv 0 (\bmod 4)$. Hence $3^\beta - 1$ is divisible by $3^4 - 1$ and so $2^{\gamma-2}5^\delta \equiv 0 (\bmod 16)$, i.e., $\gamma \geqslant 6$. Now the multiplicative order of 3 modulo $2^{\gamma-2}$ is $2^{\gamma-4}$ and the multiplicative order of 3 modulo $5^\delta$ is $4 \cdot 5^{\delta-1}$ (since 3 is a primitive root modulo 25). Hence, $\beta \equiv 0 (\bmod 2^{\gamma-4}5^{\delta-1})$ and so $\beta \geqslant 2^{\gamma-4}5^{\delta-1}$. Since $3^n > 40n$ for $n \geqslant 5$, this gives

$$3^\beta \geqslant 3^{2^{\gamma-4}5^{\delta-1}} > 2^{\gamma-1}5^\delta > 2^{\gamma-2}5^\delta + 1$$

unless $\gamma = 6$, $\delta = 1$. Thus the only solution of (A) in case 2 is $(2,4,6,1)$.

Case 3, $\alpha > 2$. Clearly $\gamma = 2$ and then the equation of (A) becomes $2^{\alpha-2}3^\beta - 1 = 5^\delta$. If $\alpha > 3$, this gives $-1 \equiv 5^\delta (\bmod 4)$, which is impossible. Thus $\alpha = 3$ and (A) becomes $2 \cdot 3^\beta - 1 = 5^\delta$. Therefore $5^{2\delta} \equiv 1 (\bmod 3^\beta)$; since 5 is a primitive root modulo 9 (and hence modulo $3^\beta$), this gives $\delta \equiv 0 (\bmod 3^{\beta-1})$. If $\beta > 1$, this gives

$$5^\delta \geqslant 5^{3^{\beta-1}} > 2 \cdot 3^\beta > 2 \cdot 3^\beta - 1,$$

While the heuristic arguments of the previous two paragraphs suggest that $r = 5$, it seems very difficult to give a rigorous proof either that $r > 2$ or that $r < 9$.

A wide variety of solutions was received. Nick Franceschine, Stephen M. Gagola, Jr., and Heinrich Rolletschek gave self-contained elementary solutions along the same lines as that given above. Gerry Myerson (Australia) gave a solution based on the elementary results of [**1**] giving all solutions of $x + y = z$ in positive integers $x$, $y$, $z$ such that $xyz$ has no prime factors larger than 7. Using [**9**] or [**10**] Kee-Wei Lau (Hong Kong) showed that there is a computable constant $N_0$ such that the assertion of the problem holds for $N \geqslant N_0$. L. E. Mattics gave a solution based on [**2**]. The proposer gave a noneffective solution similar to the argument used in [**5**, §5.13]; this is based on Thue's theorem [**3**, Theorem 116] that $ax^3 - by^3 = c$ has only finitely many integer solutions for any triple $(a, b, c)$ of positive integers. Myerson quoted the result of [**7**] that if $A, B, C, M_1, \dots, M_m, N_1, \dots, N_n$ are positive integers which are relatively prime in pairs and if $M_1, \dots, M_m, N_1, \dots, N_n$ are all greater than 1, then there are only finitely many solutions of

$$AM_1^{x_1} \cdots M_m^{x_m} - BN_1^{y_1} \cdots N_n^{y_n} = C$$

in nonnegative integers $x_1, \dots, x_m, y_1, \dots, y_n$. Finally the editors remark that the assertion of the problem can be deduced from the results of [**6**].

REFERENCES

1. Leo J. Alex, Diophantine equations related to finite groups, *Comm. Alg.*, 4 (1976) 77–100.
2. G. D. Birkhoff and H. S. Vandiver, On the integral divisors of $a^n - b^n$, *Ann. of Math.*, (2) 5 (1904) 173–180.
3. L. E. Dickson, Introduction to the Theory of Numbers, University of Chicago Press, 1929.
4. H. Halberstam and H.-E. Richert, Sieve Methods, Academic Press, 1974.
5. G. H. Hardy, Ramanujan, Cambridge University Press, 1940.
6. D. H. Lehmer, On a problem of Størmer, *Illinois J. Math.*, 8 (1964) 57–79.
7. T. Nagell, Sur une classe d'équations exponentielles, *Ark. Math.*, 3 (1958) 569–581.
8. A. Schinzel and W. Sierpínski, Sur certaines hypothèses concernant les nombres premiers, *Acta Arith.*, 4 (1958) 185–208.
9. T. N. Shorey and R. Tijdeman, Exponential Diophantine Equations, Cambridge University Press, 1986.
10. R. Tijdeman, On integers with many small prime factors, *Compositio Math.*, 26 (1973) 319–330.
11. R. C. Vaughan, A remark on the divisor function, *Glasgow Math. J.*, 14 (1973) 54–55.

## POSTMORTEM COMMENTS ON ADVANCED PROBLEMS

### Circles, Triangles, Squares, and the Golden Mean

6477 [1984, 588; 1986, 406]. *Proposed by Louis Funar, University of Craiova, Romania.*

Let $r$ be the radius of the incircle of an arbitrary triangle lying in the closed unit square. Prove or disprove that $r \leqslant (\sqrt{5} - 1)/4$.

*Editorial comment.* The solution published in 1986 appears to be incomplete. Specifically, the last inequality on line 11 from the bottom of page 406 does not always hold; this appears to be a fatal flaw in that solution. Accordingly we give the following solution here.

*Solution by Aage Bondesen, Espergaerde, Denmark.* We shall prove that $r \leqslant (\sqrt{5} - 1)/4$. By "triangle" we mean a triangle with interior. Similarly for a square. By *inflating* a triangle we mean replacing it with another triangle that contains it properly. It is not hard to show that inflating a triangle increases the inradius;

While the heuristic arguments of the previous two paragraphs suggest that $r = 5$, it seems very difficult to give a rigorous proof either that $r > 2$ or that $r < 9$.

A wide variety of solutions was received. Nick Franceschine, Stephen M. Gagola, Jr., and Heinrich Rolletschek gave self-contained elementary solutions along the same lines as that given above. Gerry Myerson (Australia) gave a solution based on the elementary results of [**1**] giving all solutions of $x + y = z$ in positive integers $x$, $y$, $z$ such that $xyz$ has no prime factors larger than 7. Using [**9**] or [**10**] Kee-Wei Lau (Hong Kong) showed that there is a computable constant $N_0$ such that the assertion of the problem holds for $N \geqslant N_0$. L. E. Mattics gave a solution based on [**2**]. The proposer gave a noneffective solution similar to the argument used in [**5**, §5.13]; this is based on Thue's theorem [**3**, Theorem 116] that $ax^3 - by^3 = c$ has only finitely many integer solutions for any triple $(a, b, c)$ of positive integers. Myerson quoted the result of [**7**] that if $A, B, C, M_1, \dots, M_m, N_1, \dots, N_n$ are positive integers which are relatively prime in pairs and if $M_1, \dots, M_m, N_1, \dots, N_n$ are all greater than 1, then there are only finitely many solutions of

$$AM_1^{x_1} \cdots M_m^{x_m} - BN_1^{y_1} \cdots N_n^{y_n} = C$$

in nonnegative integers $x_1, \dots, x_m, y_1, \dots, y_n$. Finally the editors remark that the assertion of the problem can be deduced from the results of [**6**].

REFERENCES

1. Leo J. Alex, Diophantine equations related to finite groups, *Comm. Alg.*, 4 (1976) 77–100.
2. G. D. Birkhoff and H. S. Vandiver, On the integral divisors of $a^n - b^n$, *Ann. of Math.*, (2) 5 (1904) 173–180.
3. L. E. Dickson, Introduction to the Theory of Numbers, University of Chicago Press, 1929.
4. H. Halberstam and H.-E. Richert, Sieve Methods, Academic Press, 1974.
5. G. H. Hardy, Ramanujan, Cambridge University Press, 1940.
6. D. H. Lehmer, On a problem of Størmer, *Illinois J. Math.*, 8 (1964) 57–79.
7. T. Nagell, Sur une classe d'équations exponentielles, *Ark. Math.*, 3 (1958) 569–581.
8. A. Schinzel and W. Sierpínski, Sur certaines hypothèses concernant les nombres premiers, *Acta Arith.*, 4 (1958) 185–208.
9. T. N. Shorey and R. Tijdeman, Exponential Diophantine Equations, Cambridge University Press, 1986.
10. R. Tijdeman, On integers with many small prime factors, *Compositio Math.*, 26 (1973) 319–330.
11. R. C. Vaughan, A remark on the divisor function, *Glasgow Math. J.*, 14 (1973) 54–55.

## POSTMORTEM COMMENTS ON ADVANCED PROBLEMS

### Circles, Triangles, Squares, and the Golden Mean

6477 [1984, 588; 1986, 406]. *Proposed by Louis Funar, University of Craiova, Romania.*

Let $r$ be the radius of the incircle of an arbitrary triangle lying in the closed unit square. Prove or disprove that $r \leqslant (\sqrt{5} - 1)/4$.

*Editorial comment.* The solution published in 1986 appears to be incomplete. Specifically, the last inequality on line 11 from the bottom of page 406 does not always hold; this appears to be a fatal flaw in that solution. Accordingly we give the following solution here.

*Solution by Aage Bondesen, Espergaerde, Denmark.* We shall prove that $r \leqslant (\sqrt{5} - 1)/4$. By "triangle" we mean a triangle with interior. Similarly for a square. By *inflating* a triangle we mean replacing it with another triangle that contains it properly. It is not hard to show that inflating a triangle increases the inradius;

however, for our purposes it suffices to note that inflating a triangle does not decrease the inradius. Accordingly we need only consider triangles whose vertices lie on the boundary of the closed unit square. If such a triangle has two vertices on the same edge of the unit square, we may assume that they are the endpoints of that edge; for otherwise the triangle may be inflated into one with that property. Finally, we may confine our attention to triangles at least one of whose vertices is a vertex of the unit square; for otherwise the triangle may be translated and inflated until that is the case.

Thus it suffices to consider a triangle $\Delta = OAB$ whose vertices are $O = (0,0)$, $A = (a,1)$, $B = (1,b)$, where $0 \leqslant a, b \leqslant 1$. For the purpose of comparing perimeters we introduce another triangle $\Delta' = OCD$ whose vertices are $O = (0,0)$, $C = (a+b,1)$, $D = (1,0)$. We do not exclude the possibility that $a + b > 1$. Let $x = OA = \sqrt{1+a^2}$, $y = AB = \sqrt{(1-a)^2 + (1-b)^2}$, $z = BO = \sqrt{1+b^2}$, $u = OC = \sqrt{1+(a+b)^2}$, $v = CD = \sqrt{1+(1-a-b)^2}$. Note that $u \geqslant z \geqslant 1$, $x \geqslant 1$, $v \geqslant 1$. Then

$$\begin{aligned}(u+v+1)-(x+y+z) &= \frac{u^2-x^2}{u+x} + \frac{v^2-y^2}{v+y} + \frac{1-z^2}{1+z}\\ &= \frac{2ab+b^2}{u+x} + \frac{2ab}{v+y} - \frac{b^2}{1+z}\\ &\leqslant \frac{2ab+b^2}{1+z} + \frac{2ab}{v+y} - \frac{b^2}{1+z}\\ &= 2ab\left(\frac{1}{1+z} + \frac{1}{v+y}\right)\\ &\leqslant 2ab\left(\frac{1}{2}+1\right) = 3ab\\ &\leqslant (u+v+1)ab,\end{aligned}$$

so that $(u+v+1)(1-ab) \leqslant x+y+z$ or

$$\frac{1-ab}{x+y+z} \leqslant \frac{1}{u+v+1}. \tag{$*$}$$

But

$$\frac{1-ab}{x+y+z} = \frac{2(\text{area of }\Delta)}{\text{perimeter of }\Delta} = \text{inradius of }\Delta,$$

and, on the other hand,

$$\begin{aligned}u+v+1 &= \sqrt{1+(a+b)^2} + \sqrt{1+(1-a-b)^2} + 1\\ &\geqslant \min_x\left\{\sqrt{1+x^2} + \sqrt{1+(1-x)^2} + 1\right\} = \sqrt{5}+1.\end{aligned}$$

Combining these two observations with $(*)$, we obtain

$$\text{inradius of }\Delta \leqslant \frac{1}{\sqrt{5}+1} = \frac{\sqrt{5}-1}{4} = 0.309\ldots,$$

which was to be proved.

**The Normalizer Revisited**

6528 [1986, 659; 1988, 565]. *Proposed by William P. Wardlaw, U.S. Naval Academy, Annapolis, MD.*

The following is problem 11 on page 44 of C. Chevalley's *Fundamental Concepts of Algebra*, Academic Press, 1956:

Let $H$ be a subgroup of a group $G$. Show that the elements $s$ of $G$ such that the mapping $t \to sts^{-1}$ of $G$ into itself maps $H$ into itself form a subgroup $N$ of $G$, of which $H$ is a normal subgroup; show that every subgroup of $G$ containing $H$ and in which $H$ is normal is contained in $N$.

Provide an example to show that the assertion of the problem is false.

*Correction of Previous Editorial Comment.* The 1988 editorial comment contained the following sentence [1988, 566]: "Chevalley's original assertion can be made correct either (i) by assuming that all elements of $H$ are of finite order or (ii) by defining $N$ as the set of all $s$ in $G$ such that $sHs^{-1} = H$ or (iii) by defining $N$ as the set of all $s$ in $G$ such that $sHs^{-1} \subseteq H$ and $s^{-1}Hs \subseteq H$." However, Lindsay Childs, Stephen M. Gagola, Jr., and David Manderscheid have each pointed out to the editors that this sentence is incorrect as stated but becomes correct if (i) is replaced by either

"(i)* by assuming that all elements of $G$ are of finite order"

or

"(i)** by assuming that $H$ is of finite order."

*Remark by David Manderscheid, University of Iowa, Iowa City.* The following example provides an alternative solution to the original problem and also shows that assumption (i) does not suffice to make Chevalley's assertion correct. Let $F_p$ be the field of $p$ elements for some fixed prime $p$. Let $G$ be the group of nonsingular two-by-two matrices whose entries belong to the field $F_p(t)$ of rational functions in an indeterminate $t$ over $F_p$. Let $H$ be the subgroup of $G$ consisting of the upper triangular matrices of the form

$$h = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix},$$

where $n$ belongs to the ring $F_p[t]$ of polynomials in $t$ over $F_p$. Finally let

$$s = \begin{pmatrix} t & 0 \\ 0 & 1 \end{pmatrix}.$$

Then

$$shs^{-1} = \begin{pmatrix} 1 & tn \\ 0 & 1 \end{pmatrix}, \qquad s^{-1}hs = \begin{pmatrix} 1 & t^{-1}n \\ 0 & 1 \end{pmatrix},$$

so that $sHs^{-1} \subseteq H$ but $s^{-1}Hs \not\subseteq H$. Hence the set of elements $s$ of $G$ such that $sHs^{-1} \subseteq H$ is *not* a subgroup of $G$. But note that every element of $H$ other than the identity has order $p$.

# REVIEWS

EDITED BY JOSEPH KONHAUSER
*Department of Mathematics, Macalester College, St. Paul, MN 55105*

*Derive, A Mathematical Assistant, ver. 1.22.* By Albert Rich, Joan Rich, and David Stoutemyer. Soft Warehouse, 1988. One diskette for a 512K IBM PC or NEC PC-9801 or compatibles. *Derive* manual +110 pp. Single copy: $200.

EUGENE A. HERMAN
*Grinnell College, Grinnell, IA 50112*

**Introduction**

> [T]he computer will revolutionize [mathematics] as greatly as did Arabic numerals, the invention of algebra, and the invention of calculus itself... [P]roblems solvable only as research and by an elite suddenly became routine. The computer will do the same for our mathematics, and calculus is the place to begin. [**5**, p. 173]

Sentiments of this sort, whether we agree with them or not, have been battering our consciences for several years. However, those who may have wanted to try revolutionizing mathematics teaching had little hope of doing so; almost no suitable computer software existed. Today, such a variety of potentially useful software is available that it almost defies classification. Even if we limit our attention to software suitable for the first two or three semesters of calculus, we can identify at least four categories. *Derive*, like its predecessor *muMATH*, defies easy classification; it has aspects in common with all of them.

Computer algebra systems such as *Maple* and *Mathematica* are the most ambitious packages with the broadest range of potential uses in mathematics courses. In addition to their extraordinary capabilities for symbolic computation, such as finding exact derivatives and antiderivatives of large classes of functions, they have a variety of commands for making numerical approximations and drawing graphs. *Derive* has a similar range of capabilities, including quite substantial symbolic computational power.

However, *Derive* seems especially intended for beginning calculus students, the audience targeted by such packages as True BASIC's *Calculus*, *Calculus-Pad*, and *MicroCalc*. The strength of these more specialized programs is that they attempt to cover all the computational needs of calculus students in a small package, although their capabilities for symbolic computation go no further than exact differentiation.

Now that graphics are available in some of the newest programmable scientific calculators, these too are an attractive choice for calculus students. Machines such as the *Casio fx-7055 G*, *Sharp EL-9000*, and *Hewlett-Packard 28S*, are intended to be both powerful and convenient, design goals they share with *Derive*. The HP calculator is even capable of some symbolic computation (see [**3**]).

Similar but more powerful are the packages sometimes referred to as supercalculators. (See the review [**4**].) This is not a particularly coherent category, as it includes such diverse packages as *MATLAB*, *MathCAD*, and *TKSolver Plus*. What they have in common are powerful numerical capabilities, at least some graphical capabilities,

and at least the rudiments of a programming language that lets serious users extend the package's range of capabilities. None is capable of any symbolic computation, but all run on the same class of machines as *Derive*.

How we look at *Derive* may depend largely on which category of packages we use as our frame of reference. *Derive* has graphical and numerical capabilities comparable to those found in many of the above but symbolic capabilities found in only a few. Also, only a few are designed, as *Derive* is, to be especially easy and convenient for the beginning student. But *Derive*, unlike most of the above, is not programmable and is deficient in a few other ways discussed below. So our judgement will depend on how great its strengths are, how substantial its deficiencies are, and the importance we attach to both, especially in comparison with the alternatives.

**User interface.** The *Derive* screen supports three types of windows. An "algebra" window contains a numbered list of expressions you have created; a "plot" window contains graphs of functions of one or two variables that you have selected from an algebra window; and the "authoring line" is where you create and edit expressions. Typically, you author an expression and send it to an algebra window by pressing the Enter key. Then you apply a command, such as `Integrate`, `Simplify`, or `Plot`, to the expression; this is usually done by a simple menu selection. If the result is another expression, it then shows up in the algebra window. No editing can be done in the algebra window, but you can move an existing expression to the authoring line and edit it there. Expressions are never lost, even if they scroll off the screen, as you can move the cursor back to any previous expression, and you can refer to expressions by their number when you author a new expression. You can also select any earlier expression for further action by moving to, and thereby highlighting, the expression.

These basic operations, the ones you do most often, are easy to learn and quick to carry out. More advanced operations, however, can be hard to learn and tedious. For example, to select a subexpression from an existing expression or to see parts of a long expression that has scrolled off the right side of the screen requires learning a nonintuitive use of the arrow keys with the Control key. Moving the cross hairs in a plot window and the cursor in the authoring line require yet other nonintuitive key combinations. Also, editing in the authoring line can be tedious, since you can only move its cursor by small increments from the ends of the line. These operations would be faster and more natural if *Derive*'s interface included a mouse.

Input and output of expressions is supported by a number of useful features of *Derive*. Although you enter an expression in the authoring line as a linear string of characters, much as you would when writing in a programming language, *Derive* permits many of the shorthands we employ every day when writing mathematics. Thus, it can usually interpret juxtaposition correctly as indicating multiplication and permits omission of parentheses around a single-character function argument. For example, the following expressions are legal and would be understood the way we usually intend them:

```
2(3x - 1)sin x
f(x,y) := xy
```

Understanding $xy$ as $x * y$, rather than as the variable whose name is $xy$, is possible because variable names in *Derive* are just one character long unless you make an explicit declaration to the contrary. The colon-equal combination is required in function definitions to distinguish this sign from the equal sign used in equations; however, this is a detail that beginners are shielded from if they use the menu to help them define a function. A substantial improvement of *Derive* over *muMATH* is that expressions in an algebra window appear in a readable two-dimensional format. Thus, integration, differentiation, summation, and square root symbols are used; fractions are stacked vertically; and square brackets are scaled to the height of the expression they enclose.

On the other hand, formatting is not as well done as in, for example, *MathCAD*, and a printout of the contents of an algebra window would not make a particularly useful handout. The contents of the algebra window are generated entirely by characters, not individual pixels, and so expressions always occupy an integer number of text lines. Thus, exponents are a full line higher; fractions occupy at least three lines; subscripts do not exist; and summation and square root signs do not expand vertically. Small matrices are displayed in matrix form, but a matrix that cannot fit completely on the screen is displayed linearly as a vector of (row) vectors. (See Example 7.) In the authoring line, every matrix is written as a vector of vectors. Although expressions can be rearranged in an algebra window, you cannot include plots with the expressions, and you cannot add explanatory text unless you later edit the saved file of expressions.

Sophisticated users will be disappointed by the lack of both a command language and programming language in *Derive*, although the developers have assigned such features a high priority for future versions. The closest *Derive* comes to providing a programming language construct are two commands that incorporate loops; `Sum` makes sums out of arbitrary expressions containing the summation index, and `Vector` makes vectors and matrices similarly. (See Example 9.) Also, expressions can be formed not only of arbitrary variable names and constants but of arbitrary function names, although functions cannot be passed as parameters to another user-defined function. Some, but not all, of the commands available in the menus can be included in expressions too. For example,

```
g(t) := int (f(x), x, a, t)
```

is a legitimate expression that means

$$g(t) = \int_a^t f(x)\,dx$$

provided $f$ has been declared to be a function with no value. Many of *Derive*'s commands, however, have no algebraic form and so cannot be embedded in expressions. These include `Simplify`, `Plot`, `Substitute`, `Precision`, and many more. Such deficiencies are most likely to affect instructors who wish they could expand the package's capabilities or make up special-purpose procedures for their students or prepare a detailed canned demonstration for a class. You can save expressions in a file and play them back in front of an audience, but you have to select the operations as you do so.

**Computer algebra.** *Derive*'s routines for symbolic computation comprise its most impressive collection of features. Its calculus commands produce exact derivatives,

integrals (both indefinite and definite, proper and improper), limits, Taylor polynomials, and sums and products (both finite and infinite). *Derive* can also simplify a wide variety of expressions in a number of different ways. For example, it can both factor and expand expressions and both operations can be fine-tuned via a number of options. The expressions it will simplify include polynomials (of both one and several variables), rational functions, algebraic functions, and trigonometric polynomials. Other of *Derive*'s symbolic computations will solve a polynomial equation of degree up through three (and some of higher degree) for both real and complex roots, perform the usual algebraic operations on matrices (including matrix inversion), compute the determinant and characteristic polynomial of a matrix (and hence all its real and complex eigenvalues, if its order is not over three), and row reduce a matrix.

Furthermore, *Derive*'s symbolic computation seems to be remarkably fast and efficient. The single command `Simplify` often produces exactly the simplification you want with a single keystroke. It is also flexible, since the user has access to many of the powerful simplification rules and can change how they act. For example, one can change the direction in which rules of exponents and trigonometric identifies are applied. In many ways, *Derive* is at least the equal of *Mathematica* in symbolic computation [**2**].

However, where other computer algebra systems extend the power of their symbolic computation commands by embedding them in a powerful command language and programming language, *Derive* has neither. Also, it does not have the breadth of symbolic capabilities possessed by these much larger systems. For example, it lacks matrix commands for computing nullspaces and eigenvectors, it has no differential equation solver, and its only built-in functions beyond those normally associated with beginning calculus are the error and gamma functions. Still, as a calculus package, *Derive* is missing remarkably few symbolic capabilities. Perhaps the most significant one is partial fraction decomposition, which limits integration of rational functions to those with denominators of degree one or two.

Of course, computer algebra systems are notoriously hard for college students to use [**1**], and *Derive* is not immune to this problem, despite its friendlier interface. Since the underlying algebra is complex, not real, results can take on forms that students will find unfamiliar. For example, $x^{\frac{1}{3}}$ does not plot for negative $x$, unless you change the default branch cut to be at the positive real axis rather than the negative. Also, an antiderivative may only make sense if you interpret the function as acting on a complex argument. (See Example 8.) Making substitutions and assigning values to variables in an expression is another aspect of these systems that can bewilder the beginner. In *Derive*, substituting a value for a variable and assigning a value to a variable have subtly different effects. Also, each has side effects that may confuse the user. For example, when you assign a value to a function for a particular value of its argument, the program interprets the constant value of the argument as a variable and redefines the function everywhere. Yet another confusing aspect of *Derive* is that the `Simplify` command is also used to request that an expression be evaluated, which confounds two quite different operations; separate commands would be clearer.

**Graphics.** *Derive* has two types of plot windows, 2D and 3D, and 2D plot windows are used for three different types of plots: the graph of a real-valued

function of a real variable, a polar curve, and a parametrized curve. Normally, you do not even need to specify the type of plot window you want, as *Derive* chooses a 2D or 3D plot window based on whether the expression you selected has one or two variables. Requesting a plot and specifying options for it is done much as on the HP 28S. You author and highlight an expression, select `Plot` from the algebra window menu to bring up a plot window, press a function key to change the screen from text mode to graphics mode, and select `Plot` from the plot window menu to request the plot. You could hardly ask for an easier or quicker sequence of steps. Often, you need not specify a single graphing option, not even to specify the graphing rectangle or the resolution of the plot. If the graphing rectangle, which by default is roughly the rectangle $[-4,4] \times [-2,2]$, contains the picture you want, you are done. Otherwise, you can zoom in or out by the press of a single function key which causes an expansion or contraction about the center of the plot by a factor of 2 to 2.5. Or, you can move the center and then zoom. If you see that the plot is not producing the result you wanted, you can interrupt it with a keystroke, reset some options, and try again. For example, if the resolution is poor, you can switch to a higher resolution or even to a dot mode that makes no attempt to interpolate.

Furthermore, the graphing algorithm seems remarkably robust. The program does not choke at points where the function is undefined and does not even report an error; it just plots the portions that are defined. The algorithm is also adaptive in the sense that it plots more points where the curvature of the function appears to be high. Of course, the resolution provided by the EGA board in the PC clone I use is substantially lower than that of a Macintosh II or any graphics workstation, which makes a substantial difference in the quality of the graphs. However, the quality of *Derive*'s graphs is certainly far superior to that found in graphing calculators.

On the other hand, *Derive* is missing some useful plotting options. For example, you cannot specify the bounds of the rectangular 2D plot region. If you want to set these bounds exactly, you must translate the center of the plot and multiply the horizontal and vertical scales by factors that are not always easy to compute. Furthermore, if you have found reasonable horizontal bounds but do not know what vertical bounds will include the interesting part of the graph, you cannot ask for automatic scaling in the vertical direction. Although you can plot functions, you cannot plot data points, which is part of a broader class of missing features discussed in the next section. You can graph a function of two variables in a 3D plot but not its level curves or its gradient field. You can get a screen dump of a plot, but there is no command that lets you send a higher resolution plot to a laser printer, for example.

The quality of 3D plots in *Derive* is good, although it does not rival *Mathematica's* [**2**]. *Derive*'s 3D plots do scale automatically in the vertical direction, and they use different colors to distinguish between the upper and lower sides of the surface. The speed of plotting, even with hidden lines removed, is reasonable: 15 seconds for a typical plot using a 10 by 10 grid (the default) and 35 seconds using a 20 by 20 grid. However, the graphs lack nice touches such as surface shading and lighting, and every change of the options requires a complete redrawing of the graph.

**Numerical approximations.** The numerical capabilities of *Derive* are its least impressive. It does have the most essential ones for calculus: numerical integration for single and multiple integrals, and numerical solution of a single nonlinear

equation and a system of linear equations. Also, for business calculus courses, there are built-in financial functions for computing present and future values and periodic payments. But again, the greatest strength of *Derive* is that it lets you do routine operations quickly and easily. Typically, you select an expression, such as a definite integral or any expression that evaluates to a constant, select `approX` from the menu, and you have your answer accurate to about six decimal places. If you are planning a number of approximations or want more places of accuracy, you use the menus to change the default mode of precision from exact to approximate or increase the number of digits of precision. A nice advantage of combining symbolic and numeric computation in one package is that each can support the other. Thus, since *Derive*'s real arithmetic computations are carried out in software, you can have any number of digits of precision. Also, you can sometimes avoid pitfalls of numeric computation, such as cancellation errors and slow convergence, by computing a result exactly and then approximating this exact result. (See Examples 4 and 10.)

However, *Derive* is missing quite a bit compared to the supercalculators mentioned earlier. For use in calculus, its major omission is commands for manipulating numeric data. You cannot plot data points or generate a sequence of approximations or perform a least-squares fit. Although there are functions for computing an average and variance, these functions require explicit arguments and cannot act on data in a vector or matrix, the natural places to store data. Another numerical capability that would be useful in calculus but is absent from *Derive* is solution of a system of nonlinear equations. Also not provided are commands for finding complex roots of polynomial equations (and hence finding complex eigenvalues), matrix computations other than matrix inversion and computation of real eigenvalues from the characteristic polynomial, and commands for solving differential equations. A concern for instructors who may wish their students to learn from numerical approximations is that *Derive*'s commands simply produce a single numerical result with no information about its reliability other than an occasional message reporting that the result has "dubious accuracy." No intermediate results are available and little information is provided about the methods used.

The speed of *Derive*'s numeric computations is also inferior to that found in the supercalculators. (See [4] and Example 10.) This may be largely a result of its real arithmetic being carried out in software.

**Documentation.** The *Derive* manual is only adequate. The explanations of the package's features are clear and the index is satisfactory, but there is no reference section containing an authoritative summary of each command. In fact, a few commands, such as `Solve`, are nowhere completely documented. Another serious omission is any discussion of the program's data structures or any other attempt to explain more deeply how functions and variables and expressions are likely to interact with one another or how they are affected by substitutions and assignments. The on-line help is minimal.

The worked examples in the manual are almost all simple, single-line commands, as are the examples on the diskette. Those who want to use the package in less routine ways would benefit from a few longer, more sophisticated examples.

**Examples.** The examples below are all taken or adapted from the reviews [**2**, **3**, **4**] to provide a modest basis of comparison. I did them on a *Compaq Deskpro 286* with

640K of memory and a 80287 math coprocessor, although *Derive* makes no use of the coprocessor. However, I present the solutions with less use of menus than a student's solution would, since menu selections would be hard to follow in print.

1. *Integrate*. Evaluate symbolically and check by differentiation

$$\int x^4\sqrt{a^2 - x^2}\, dx.$$

The following two expressions produced the integral symbolically (in 1.1 seconds), then differentiated the result and simplified the derivative to its original form (in 2.3 seconds). The "#1" refers to the first expression, which was numbered 1 in the algebra window.

```
int(x^4√(a^2 - x^2), x)
dif(#1,x)
```

Using menus instead, one would author only the integrand.

2. *Taylor*. Evaluate symbolically the Taylor polynomial of degree seven about 0 for

$$\int_0^x \sqrt{1 - t^3}\, dt.$$

The first step below simply created an expression in the algebra window that resembled the integral as displayed above; the second produced the Taylor polynomial (in 3.0 seconds).

```
f(x) := int(√(1 - t^3), t, 0, x)
taylor(f(x), x, 0, 7)
```

However, the constant term was left as $\int_0^0\sqrt{1 - t^3}\, dt$ and would not simplify.

3. *Lissajous*. Graph $x = \sin(2t + 0.6)$, $y = \sin(5t)$ as $t$ ranges over $[0, 2\pi]$.

I authored the expression `[sin(2t + 0.6), sin(5t)]`, then selected `Plot` from the menu, pressed a key to switch to graphics mode, selected `Plot` from plot menu, pressed `Enter` to accept the default $t$ range of $[-\pi, \pi]$, and got the graph in 16 seconds.

4. *Cancellation error*. Evaluate, for $x = 0.707107$,

$$f(x) = 8118x^4 - 11482x^3 + x^2 + 5741x - 2030.$$

I first evaluated $f(0.707107)$ exactly and selected `approX` from the menu to get the accurate result $-1.91527\ 10^{-11}$. Then I changed the precision to 15 digits, selected `approX` again, and got the accurate result $-1.91527325270819\ 10^{-11}$. However, when I changed the mode of precision to approximate rather than exact and recomputed $f(0.707107)$, I got 0.

5. *Roots*. Find the roots of $x^3 - x^2 - x - 1 = 0$.

```
solve(x^3 - x^2 - x - 1, x)
```

produced the two complex roots and one real root exactly (in 4.5 seconds). To get decimal approximations, I highlighted the collection of roots and selected `approX`. Numerically, however, I could only compute the real root, which I did with

```
solve(x^3 - x^2 - x - 1, x, - 10, 10)
```

6. *Prime factors*. Decompose 266382003787 into its prime factors.

After entering the number, I selected **Factor** from the menu and got the prime factors 151667 and 1756361 (in 3.6 seconds). I was so encouraged that I had *Derive* factor 2098305176053226299 into its prime factors 530110817 and 2958238747, which it did in 2.7 minutes.

7. *Eigenvectors*. Find the eigenvalues and eigenvectors of $\begin{bmatrix} 6 & 4 & 4 & 1 \\ 4 & 6 & 1 & 4 \\ 4 & 1 & 6 & 4 \\ 1 & 4 & 4 & 6 \end{bmatrix}$

Although *Derive* has no command for computing eigenvectors, it was able to do most of the work as follows. I authored the matrices $A$ and $xI - A$ by

```
a:= [[6,4,4,1], [4,6,1,4], [4,1,6,4], [1,4,4,6]]
c(x) := x identity_matrix(4) - a
```

The command `eigenvalues(a)` produced `[w = - 1, w = 5, w = 15]`, and the command `row_reduce(c( - 1))` (plus two more like it), allowed me to quickly spot bases for the three eigenspaces.

8. *Differential equation*. Solve numerically and symbolically the following initial-value problem and graph the solution:

$$y' = 1 - 2xy,\ y(-1) = 0, \text{on } [-1,2].$$

The first expression below is the command I gave to produce the exact solution; the second is the result.

```
int(exp(x^2), x, - 1, t)
```

$$\mathrm{i}\left[-\frac{\sqrt{\ }(\pi)\mathrm{erf}(\mathrm{i})}{2} - \frac{\sqrt{\ }(\pi)\mathrm{erf}(\mathrm{it})}{2}\right]$$

This is indeed correct and real-valued, but it would not plot since *Derive* cannot evaluate the error function numerically at complex arguments. No numerical procedure is available for solving a differential equation, and the user cannot write such a procedure.

9. *Series solution*. Find the terms through $x^6$ of the power series solution of the initial-value problem

$$y'' + \sin(x)y = 0,\ y(0) = 0,\ y'(0) = 1.$$

Although the *Derive* manual makes no mention of subscripted variables, one can simulate them by creating an undefined function which is then applied to integer arguments.

```
a(x) :=
sum(a(k)x^k / k!, k, 0, 6)
```

The second line creates a general sixth degree polynomial with the undetermined coefficients $a_k/k!$ Then I substituted 0 for $a(0)$ and 1 for $a(1)$ by highlighting the two expressions to be replaced and selecting **Substitute** from the menu. The following steps complete the solution. (However, this is a nonstandard use of *Derive*

and may be trickier to make work in other versions of the program. For example, it seems necessary to **Simplify** after each step.)

```
f(x) := #4
g(x) := dif(f(x), x, 2)+sin(x) f(x)
h(x) := taylor(g(x), x, 0, 4)
s := vector(a(j), j, 2, 6)
e := vector(h(j), j, - 2, 2)
solve (e, s)
```

These six steps have the following results. First, $f(x)$ is defined to be our fourth expression, which is the sixth degree polynomial after the two substitutions. $g(x)$ is the left side of the differential equation, and $h(x)$ is its Taylor polynomial of degree 4. $s$ is the vector of unknowns $[a(2),\dots,a(6)]$, and $e$ is a suitable vector of equations, $[h(-2), h(-1), h(0), h(1), h(2)]$ (where "$=0$" is understood). The last line solves the five linear equations for the five unknowns and produces

```
[a(2) = 0  a(3) = 0  a(4) = - 2  a(5) = 0  a(6) = 4].
```

Finding no easy way to substitute this back into the polynomial $f(x)$, I simply wrote the result on paper:

$$f(x) = x - \frac{1}{12}x^4 + \frac{1}{180}x^6.$$

10. *Gibbs phenomenon.* The $n$th partial sum of the Fourier series for $(\pi/4)\,\mathrm{sgn}(x)$ is

$$f_n(x) = \sum_{k=1}^{n} \frac{\sin(2k-1)x}{2k-1}.$$

Study the Gibbs overshoot graphically and analytically, emphasizing $f_{20}$. In particular, graph $f_{20}$ on the interval $[0, \pi/2]$, find its maximum on this interval, and find its average value on $[1,2]$ by computing $\int_1^2 f_{20}$. Also, compare the graphs of $f_8$, $f_9$, $f_{10}$ on the same set of axes, and evaluate $f_{500}(x)$ and $f_{10000}(x)$ for a few values of $x$.

First I authored

```
f(x,n) := sum(sin((2k - 1)x)/(2k - 1), k, 1, n).
```

Plotting $f_{20}(x) = f(x,20)$ took over $1\frac{1}{2}$ minutes, probably because the wiggly graph caused the adaptive graphing algorithm to plot many points. Centering and scaling took even longer, since I did not get the graphing rectangle I wanted on my first two tries. I then used the plot window's cross hairs to get a rough approximation to the coordinates of the maximum point. Next, I found a better approximation by setting the derivative of $f_{20}$ equal to zero and solving numerically in the small interval I had located graphically.

```
g(x) := dif(f(x, 20), x)
solve (g(x), x, 0.5, 0.1)
```

The solution, accurate to 6 digits, was $x = 0.07854$ (in 22 seconds); the value of $f(0.07854)$ was 0.926131. Using 12 digits of precision, the solution took 92 seconds. Computing $\int_1^2 f(x,20)\,dx$ numerically proved lengthy; I aborted the computation after five minutes. But I computed its exact value (in 13.4 seconds) and approxi-

mated this result numerically (in 4.1 seconds) using 12 digits of precision. The quite accurate result was 0.78602335388.

Next I placed expressions for $f_8$, $f_9$, and $f_{10}$ in the algebra window and created a side-by-side plot window. So I was able to select an expression, switch to the plot window and plot it, switch back to the algebra window, and repeat. The three curves were plotted in three different colors, without my having requested this, and they were easily distinguishable. Evaluating $f(1,500)$ numerically, using 6 digits of precision, took 20.6 seconds. Evaluating $f(1,10000)$ took over six minutes, still using 6 digits.

**Conclusion.** How you judge *Derive* will depend greatly on your needs and your resources. As a student-oriented calculus package for the IBM PC and its relatives, *Derive* is extraordinarily complete and easy to use. Routine operations can be carried out by beginners easily and quickly, and the operations that are routine in this package are often not even available in others. The most crucial part of any calculus package is its graphics, and *Derive*'s graphics are indeed worth having even though a PC won't fit in your purse or pocket. The numeric capabilities are next most important, and here the developers have more work to do; commands that act on data are especially needed. The symbolic computations, though we may be uncertain how they will benefit calculus students, are a great strength of the package and support its other capabilities. The user interface is indeed much easier to use than that of other computer algebra systems, but mouse enthusiasts will find *Derive*'s cursor movements tedious and awkward. Also, calculus instructors will find their use of the package limited by its lack of programmability and its unsuitability for producing attractive handouts. Much handier would be a graphical interface that permits expressions, plots, and text to be freely intermixed.

*Derive*'s coverage does not range much beyond calculus. It has some useful matrix commands, but they are not yet sufficient to constitute a linear algebra package. Of course, *Derive* is still under development, and new versions are coming out in rapid succession. We can expect much more yet, as the developers are quite experienced, having started their company and *muMATH* in 1979. However, I look forward to seeing whether they can succeed at adding programmability to the package without losing its simplicity. No computer algebra system has yet succeeded in presenting a conceptually simple model of itself that lets the inexperienced user harness its power after a brief learning time. *Derive* is no exception; sophisticated, multi-step operations can be tricky, and the manual offers no conceptual model of how the underlying algebra works.

For a new package, *Derive* is remarkably bug free. Of the few bugs I reported, almost all had been noted and corrected. And in all my sessions with the package, *Derive* never crashed or required rebooting.

**Other Packages Mentioned**

*Calculus*, True BASIC, Hanover, NH.
*Calculus-Pad*, Brooks/Cole, Pacific Grove, CA.
*Maple*, WATCOM, Waterloo, Ontario, Canada, and Brooks/Cole, Pacific Grove, CA.
*MathCAD*, MathSoft, Cambridge, MA.
*Mathematica*, Wolfram Research, Champaign, IL.

*MATLAB*, Math Works, South Natick, MA.
*MicroCalc*, MathCalcEduc, Ann Arbor, MI.
*muMATH*, The Soft Warehouse, Honolulu, HI.
*TKSolver Plus*, Universal Technical Systems, Rockford, IL.

REFERENCES

1. M. Henle, Another calculus course with computers: discrete and continuous modeling (Preprint).
2. E. A. Herman, Mathematica—a review, *Notices of the AMS*, 35 (1988) 1334–1344.
3. Y. Nievergelt, The chip with the college education: the HP-28C, this MONTHLY, 94 (1987) 895–902.
4. B. Simon and R. M. Wilson, Supercalculators on the PC, *Notices of the AMS*, 35 (1988) 978–1001.
5. G. S. Young, Present problems and future prospects, Calculus for a New Century: A Pump, not a Filter, MAA Notes Number 8, 1987, pp. 172–175.

# TELEGRAPHIC REVIEWS

*Edited by Lynn Arthur Steen, with the assistance of the Mathematics Departments of Carleton, Macalester, and St. Olaf Colleges. Books and software submitted for review should be sent to Reviews Editor, American Mathematical Monthly, St. Olaf College, Northfield, Minnesota 55057.*

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

| | | |
|---|---|---|
| T: Textbook | P: Professional Reading | 1–4: Semesters |
| C: Computer Software | L: Undergraduate Library | **: Special Emphasis |
| S: Supplementary Reading | 13: Grade Level | ??: Questionable |

Readers are advised that price information is subject to change, that computer software is often available also on other machines, and that hardware variations often cause software incompatibilities. Selected books and software packages receive a second, more extensive review in the MONTHLY.

**General, S, L*.** *The Privilege of Being a Physicist.* Victor F. Weisskopf. WH Freeman, 1989, x + 235 pp, $17.95. [ISBN: 0-7167-1982-7] A collection of sixteen essays from the last fifteen years, revised to suit current circumstances, on science and culture (education, art, limits of science); ideas in physics (quantum mechanics, elementary particles, origin of the universe); physicists (Pauli, Heisenberg); science and society (peace, Hitler, technology, nuclear bombs); and the life of a scientist (privileges, threats, promises). The essays return again and again to Weisskopf's main concern: nuclear arms, ethics, and responsibility. LAS

**Precalculus, T(13).** *College Algebra, Fourth Edition.* Karl J. Smith, Patrick J. Boyle. Brooks/Cole, 1989, xii + 579 pp, $37.25. [ISBN: 0-534-09504-6] Standard algebra text for precalculus that does include exponential and logarithmic functions, the binomial theorem, combinatorics and probability, and a chapter on mathematical proof. (*Second Edition*, TR, January 1983; *Third Edition*, TR, November 1985.) AWR

**Education, P, L.** *Setting a Research Agenda.* Ed: Judith T. Sowder. Res. Agenda for Math. Educ., V. 5. Lawrence Erlbaum Associates & NCTM, 1989, v + 64 pp, $7.50 (P). [ISBN: 0-87353-269-4] Brief concluding volume in the NCTM Research Agenda Series prepared by the editorial committee as a backdrop to the preceding four volumes (TR's, June-July 1988; January 1989; October 1989; November 1989). Focuses on research and reform in school mathematics, on the nature of mathematical knowledge, and on the process of instruction; emphasizes opportunities for research that responds to issues that confront mathematics education. LAS

**History, S*(13-16), P*, L**.** *Who Got Einstein's Office?* Ed Regis. Addison-Wesley, 1987, xv + 316 pp, $10.95 (P). [ISBN: 0-201-12278-2] A compelling account of people and ideas at the Institute for Advanced Study—Einstein, Gödel, von Neumann, Oppenheimer, Wolfram; string theory, fractals, cellular automata, and more. Superb scientific journalism that tells the story of uncompromising creativity and imagination in the pursuit of science. Thoroughly researched and well-documented, the book nonetheless reads like a spirited tale of intellectual intrigue. LAS

**Foundations, T(14-17: 1, 2), S, L.** *Computability: Computable Functions, Logic, and the Foundations of Mathematics.* Richard L. Epstein, Walter A. Carnielli. Math. Ser. Wadsworth, 1989, xvii + 297 pp, $44.95. [ISBN: 0-534-10356-1] Part one presents philosophical background and basic notions of whole number, real number, function, proof, while the second part describes computability, Turing machines, partially recursive functions. Part three deals with propositional and predicate logic. Part four is devoted to more advanced ideas including Church's Thesis and constructivity. Lots of historical information and quotes. Exercises, bibliography, glossary, index. RJA

**Graph Theory, P.** *Distance-Regular Graphs.* A.E. Brouwer, A.M. Cohen, A. Neumaier. Ser. of Mod. Surveys in Math., B. 18. Springer-Verlag, 1989, xvii + 495 pp, $84.50. [ISBN: 0-387-50619-5] Main emphasis of book is to describe, classify, and characterize the known distance-regular graphs. Discusses association schemes, $Q$-polynomial distance-regular graphs, distance-transitive graphs, graphs of Coxeter and Lie types, graphs related to codes and classical geometries, and representations of graphs. Comprehensive book with 800 references. Not intended as an introduction to algebraic graph theory. LC

**Combinatorics, P, L.** *Introduction to Arrangements.* Peter Orlik. CBMS Reg. Conf. Ser. in Math., No. 72. AMS, 1989, x + 110 pp, (P). [ISBN: 0-8218-0723-4] An *arrangement* is a finite set of hyperplanes of codimension in some vector space. This monograph introduces the combinatorics, cohomology, topology, and geometry of arrangements, building on and pulling together a vast world-wide literature that treats various special cases. Based on expository lectures from a June 1988 CBMS Regional Research Conference at Northern Arizona University. LAS

**Combinatorics, P, L*.** *Combinatorial Algorithms:*

*An Update.* Herbert S. Wilf. CBMS-NSF Reg. Conf. Ser. in Appl. Math., V. 55. SIAM, 1989, v + 47 pp, (P). [ISBN: 0-89871-231-9] Lecture notes from a 1987 conference held at Colorado College providing recent research results on various coding and tree algorithms to update the 1978 book *Combinatorial Algorithms* by Nijenhuis and Wilf (TR, June-July 1979). Specific topics include Gray codes, rooted trees, free trees, and random graphs. LAS

**Combinatorics, P.** *Introduction to Coding Theory and Algebraic Geometry.* Jacobus H. van Lint, Gerard van der Geer. DMV Seminar, Band 12. Birkhauser Boston, 1988, 83 pp, $18.50 (P). [ISBN: 0-8176-2230-6] Based on lecture notes given at Schloss Mickeln, Düsseldorf, November 1987. Introduction to coding theory and the ideas of algebraic geometry which led Tsfasman, Vladut, and Zink to their construction of codes which exceed the Gilbert-Varshamov bound. LC

**Number Theory, P.** *Eight Papers Translated from the Russian.* V.A. Belonogov, *et al.* AMS Transl. Ser. 2, V. 143. AMS, 1989, vii + 152 pp, $58. [ISBN: 0-8218-3124-0] One major expository paper by Zarkhin and Parshin on Falting's work occupies half of this slim volume. Other short papers are on aspects of logic, number theory, and algebra. LAS

**Number Theory, P.** *The Markoff and Lagrange Spectra.* Thomas W. Cusick, Mary E. Flahive. Math. Surveys & Mono., No. 30. AMS, 1989, ix + 97 pp, $42. [ISBN: 0-8218-1531-8] An exposition of modern results and modern proofs of classical results concerning the Markoff and Lagrange spectra. The authors present theorems comparing the spectra, discussing gaps in the spectra, and comparing measures of subsets of the sets. SG

**Linear Algebra, S(13), L.** *Vector Algebra: Worked Examples with Elements of Theory.* P. Gusyatnikov, S. Reznichenko. MIR (US Distr: Imported Pub), 1988, 288 pp, $8.95 (P). [ISBN: 5-03-000215-4] Presents various methods for solving geometrical problems using vector algebra and analytic geometry. Revised from 1985 Russian edition. JNC

**Group Theory, T(18), P.** *Lectures on Buildings.* Mark Ronan. Perspect. in Math., V. 7. Academic Pr, 1989, xii + 201 pp, $27.95. [ISBN: 0-12-594750-X] Notes from a graduate course which expose the basic theory of buildings beginning with chamber systems and Coxeter complexes and including discussions of spherical and affine buildings. SG

**Group Theory, S(18), P.** *Representation Theory and Harmonic Analysis on Semisimple Lie Groups.* Ed: Paul J. Sally, Jr., David A. Vogan, Jr. Math. Survey & Mono., No. 31. AMS, 1989, xi + 350 pp, $88. [ISBN: 0-8218-1526-1] A collection of five papers, all at least fifteen years old but not previously published. Three are theses by J. Arthur, M.S. Osborne, and W. Schmid; the other two by R. Langlands and P. Trombi. Each presents important ideas related to representation theory and harmonic analysis. JS

**Group Theory, S(18), P.** *Algebraic Theory of the Bianchi Groups.* Benjamin Fine. Pure & Appl. Math., V. 129. Marcel Dekker, 1989, vii + 249 pp, $99.75. [ISBN: 0-8247-8192-9] The two major algebraic results presented here are a decomposition of Bianchi groups as a non-trivial free product with amalgamation and a demonstration of the existence of subgroups which are non-congruence subgroups. Additional results for the Modular group, Euclidean Bianchi groups, the Picard group, and Fuchsian subgroups. References, index. JS

**Algebra, S(16-18), P, L.** *Infinite Algebraic Extensions of Finite Fields.* Joel V. Brawley, George E. Schnibben. Contemp. Math., V. 95. AMS, 1989, xv + 104 pp, $24 (P). [ISBN: 0-8218-5101-2] Building on work by Steinitz in 1910 on infinite algebraic extensions, the authors are able to extend many of the classical results on finite fields. Written so as to be accessible to an advanced undergraduate, the exposition begins with a survey of finite fields and continues to iterated presentations, explicit bases, polynomials, and extended polynomials. Applications to Latin squares and permutation polynomials. Bibliography. JS

**Algebra, S(17-18), P.** *Lecture Notes in Mathematics-1371: Commutative Coherent Rings.* Sarah Glaz. Springer-Verlag, 1989, xi + 347 pp, $36.30 (P). [ISBN: 0-387-51115-6] A ring $R$ is *coherent* if every finitely generated ideal of $R$ is finitely presented (e.g., Noetherian rings). Part of the study historically consisted of asking which "Noetherian" results transferred to the "coherent" case. This book covers most known results in commutative coherent ring theory including previously unpublished results, blending the two approaches to the subject: the ring theoretic and homological viewpoints. With each of the main topics, an historical overview is provided. Assumes basic commutative and homological algebra, and includes a chapter on preliminaries. SB

**Algebra, S(16), L.** *Théorie des corps: La règle et le compas.* Jean-Claude Carrega. Hermann, 1989, xii + 277 pp, 136 F (P). [ISBN: 2-7056-1402-8] An interesting introduction to field theory via the classical construction problems of geometry. Contains many historical remarks and passages. Unfortunately, the book does not fit easily into the American undergraduate curriculum. SG

**Algebra, S(18), P.** *Lecture Notes in Mathematics-1373: Lie Algebras, Madison 1987.* Ed: Georgia Benkart, J. Marshall Osborn. Springer-Verlag, 1989, 143 pp, $16.20 (P). [ISBN: 0-387-51147-4] Proceedings of a workshop on Lie algebras held in Madison, Wisconsin in August 1987. Nine papers, including major work by H. Strade, R.L. Wilson. List of lectures and participants. JS

**Calculus, S(13), L*.** *Prof. E. McSquared's Expanded Intergalactic Version: A Calculus Primer.* Howard Swann, John Johnson. Janson, 1989, 261 pp, $15.50 (P). [ISBN: 0-939765-12-8] New edition (with added exercises) of a contemporary classic (TR, June-July 1975). What Silvanus Thompson's *Calculus Made Easy* did for an earlier generation,

this fantasized primer on calculus does for today's frightened students: it de-mystifies calculus through cartoon characters—wild, funny, but ever-truthful. You *can* learn important parts of elementary calculus from this book, but you can also read it for just plain fun. LAS

**Real Analysis, T(13-14), L.** *Mathematical Analysis in Questions and Problems.* Ed: B.F. Butuzov. MIR (US Distr: Imported Pub), 1988, 271 pp, $8.95 (P). [ISBN: 5-03-000217-0] A fine source of problems for an honors calculus or first course in real variables. Differs from books of problems in that each topic is given four sections: summary of theory and formulas (no proofs); simple problems to check basic comprehension; worked problems—including some selected from summary above; problems and exercises for independent work. Intended as a companion volume for a standard text at Moscow University. AWR

**Complex Analysis, P.** *Analyticity in Infinite Dimensional Spaces.* Michel Hervé. Stud. in Math., V. 10. Walter de Gruyter, 1989, viii + 206 pp, $64.95. [ISBN: 0-89925-205-2] An advanced expository monograph on the theory of analytic functions—in a suitably generalized sense—between (possibly) infinite-dimensional complex vector spaces. After three introductory chapters, attention centers on plurisubharmonicity and related issues, e.g., the (still open) infinite-dimensional analogue of the classical Levi problem. With brief chapter summaries, a few exercises, and an extensive bibliography. PZ

**Differential Equations, T(15: 1), C.** *Elementary Differential Equations with Applications, Second Edition.* C.H. Edwards, Jr., David E. Penney. Prentice-Hall, 1989, xvi + 567 pp. [ISBN: 0-13-254046-0] A revised and reduced version of the 665-page *First Edition* (TR, October 1985). Gone (to another version of the book—see following review) are the chapters on Fourier series and boundary value problems, while the chapter on numerical methods has been extensively rewritten. Illustrative Basic programs included, with Basic, Fortran, Pascal, and APL versions available on diskette. SP

**Differential Equations, T(14: 1, 2), C.** *Elementary Differential Equations with Boundary Value Problems, Second Edition.* C.H. Edwards, Jr., David E. Penney. Prentice-Hall, 1989, xvii + 696 pp [ISBN: 0-13-254061-4]; *Solutions Manual,* vii + 338 pp, (P). [ISBN: 0-13-254079-7] An expanded version of *Elementary Differential Equations with Applications, Second Edition,* by the same authors. In addition to the chapters on ordinary differential equations found in the shorter work, this book also has chapters on Fourier series and boundary value problems. Numerical methods are illustrated with Basic programs in the text. A diskette with the same procedures coded in Basic, Fortran, Pascal, and APL is available from the publisher. Approximately 1800 exercises, mostly computational in nature. A few purely theoretical exercises are available. Prerequisite is a year of calculus. Includes the linear algebra necessary for linear systems of differential equations. SM

**Differential Equations, T(18), P.** *Symmetries and Differential Equations.* George W. Bluman, Sukeyuki Kumei. Appl. Math. Sci., V. 81. Springer-Verlag, 1989, xiii + 412 pp, $54.[ISBN: 0-387-96996-9] Notable additions to previous similar books include chapters on dimensional analysis, potential symmetries, and mappings between differential equations, as well as a good overview of applications of symmetry groups to ordinary differential equations. Lots of algorithms and exercises, plus attention to boundary value problems where relevant. JO

**Differential Equations, T(16-17: 1), S, P.** *Local Methods in Nonlinear Differential Equations.* Alexander D. Bruno. Transl: William Hovingh, Courtney S. Coleman. Springer-Verlag, 1989, x + 348 pp, $119. [ISBN: 0-387-18926-2] Develops the method of normal forms to study the local properties of systems of nonlinear ordinary differential equations. Preliminary analytical techniques presented in first chapters. Traditional applications to ordinary differential equations, mechanics, and bifurcation theory follow. SP

**Numerical Analysis, P.** *Nonlinear Numerical Methods and Rational Approximation.* Ed: Annie Cuyt. Math. & Its Applic. D Reidel (US Distr: Kluwer Academic), 1988, xviii + 458 pp, $99. [ISBN: 90-277-2669-8] Proceedings of a conference held at the University of Antwerp, Belgium in April 1987. Subfields in which papers were presented: Padé approximation and rational interpolation, rational approximation, multidimensional and multivariate problems, orthogonal polynomials and the moment problem, continued fractions, and applications. AWR

**Operator Theory, T(18: 1), P.** *Operator Theory and Arithmetic in $H^\infty$.* Hari Bercovici. Math. Surv. & Mono., No. 26. AMS, 1988, xii + 275 pp, $67. [ISBN: 0-8218-1528-8] The class $C_0$ consists of contraction operators for which the associated functional calculus on $H^\infty$ has a nontrivial kernel. This class admits a classification theorem analagous to Jordan's classification theorem for linear transformations on a finite-dimensional vector space, and is the central object of study in this monograph. Prerequisite is a course in functional analysis at the graduate level. AWR

**Operator Theory, S(18), P.** *Coxeter Graphs and Towers of Algebras.* Frederick M. Goodman, Pierre de la Harpe, Vaughan F.R. Jones. Math. Sci. Res. Inst. Publ., V. 14. Springer-Verlag, 1989, x + 288 pp, $32. [ISBN: 0-387-96979-9] After an initial chapter dealing with matrices with integral entries and relations to Coxeter graphs, the next two chapters are based on recent works of V.F.R. Jones pertaining to constructions of towers of subalgebras for multi-matrix algebras and finite von Neumann algebras. Chapter 4 considers commuting squares and derived towers. Appendices, bibliography, index. JS

**Functional Analysis, S(18), P.** *J Contractive Matrix Functions, Reproducing Kernel Hilbert Spaces and Interpolation.* Harry Dym. CBMS

Reg. Conf. Ser. in Math., No. 71. AMS, 1989, x + 147 pp, $22 (P). [ISBN: 0-8218-0722-6] An expanded and updated version of lectures from 1984 on $J$ inner matrices. Topics include reproducing kernel Hilbert spaces, the Nevanlinna-Pick problem, and interpolations. Appendix, references. JS

**Functional Analysis, P.** ***Lecture Notes in Mathematics-1376: Geometric Aspects of Functional Analysis.*** Ed: J. Lindenstrauss, V.D. Milman. Springer-Verlag, 1989, vii + 288 pp, $24.30 (P). [ISBN: 0-387-51303-5] Proceedings of a seminar in Israel. Focus is on the geometry of convex bodies, Banach space theory, and the connection between them. The majority of papers report on original research. AWR

**Functional Analysis, P.** ***Measure and Measurable Dynamics.*** Ed: R. Daniel Mauldin, R.M. Shortt, Cesar E. Silva. Contemp. Math., V. 94. AMS, 1989, xxiii + 326 pp, $48 (P). [ISBN: 0-8218-5099-7] Proceedings of a conference in honor of Dorothy Maharam Stone held in September 1987 at the University of Rochester. AWR

**Functional Analysis, T(17-18: 3), S, P, L.** ***Linear Integral Equations.*** Rainer Kress. Appl. Math. Sci., V. 82. Springer-Verlag, 1989, xi + 299 pp, $49. [ISBN: 0-387-50616-0] A well-balanced introduction to integral equations. Reviews necessary functional analysis, including normed spaces, bounded and compact operators, and Riesz-Fredholm theory. Continues with classical applications such as the Laplace and heat equations. Second half of book investigates numerical solution methods and ill-posed equations in a Hilbert space setting. KS

**Functional Analysis, P.** ***Cordes' Two-Parameter Spectral Representation Theory.*** D.F. McGhee, R.H. Picard. Pitman Res. Notes in Math. Ser., V. 177. Longman Scientific and Technical (US Distr: Wiley), 1988, 114 pp, $41.95 (P). [ISBN: 0-470-21084-2] Multiparameter spectral theory generalizes the familiar spectral representation theory of operators in separable Hilbert spaces. Its results have implications for separation-of-variables methods for solving partial differential equations. The present work, which treats mainly the two-parameter case, draws on and up-dates results obtained in the 1950's by H.O. Cordes. PZ

**Functional Analysis, P.** ***Harmonic Analysis of Spherical Functions on Real Reductive Groups.*** Ramesh Gangolli, V.S. Varadarajan. Ergebnisse der Math. und ihrer Grenzgebiete, V. 101. Springer-Verlag, 1988, xiv + 365 pp, $110. [ISBN: 0-387-18302-7] A self-contained account of the spectral theory of symmetric spaces. Presents spherical functions, semi-simple Lie groups, Harish-Chandra series, asymptotics, $L^p$ theory, and Fourier analysis on symmetric spaces. SP

**Analysis, T(18: 1).** ***Large Deviations.*** Jean-Dominique Deuschel, Daniel W. Stroock. Pure & Appl. Math., V. 137. Academic Pr, 1989, xiv + 307 pp, $34.95. [ISBN: 0-12-213150-9] Surveys a few outstanding successes in the study of large deviations and attempts to identify the common principles which support these successes. Intended as a text for advanced graduate students with a strong background in analysis and some knowledge of probability. Exercises are included after each section. SM

**Analysis, P.** ***Analysis at Urbana, Volume II: Analysis in Abstract Spaces.*** Ed: E. Berkson, T. Peck, J. Uhl. London Math. Soc. Lect. Note Ser., V. 138. Cambridge U Pr, 1989, 356 pp, $34.50 (P). [ISBN: 0-521-36437-X] Proceedings of the special year in modern analysis held in 1986-87 at the University of Illinois. *Volume I* covered analysis in function spaces. AWR

**Algebraic Geometry, P.** ***Lecture Notes in Mathematics-1378: Capacity Theory on Algebraic Curves.*** Robert S. Rumely. Springer-Verlag, 1989, 437 pp, $37.30 (P). [ISBN: 0-387-51410-4] Using potential theory as a model, the author develops a capacity theory for adelic sets on algebraic curves. His principal result, concerning curves over global fields, is the analogue of a classical theorem of Fekete and Szegö. SG

**Geometry, P, L.** ***Elementary Geometry in Hyperbolic Space.*** Werner Fenchel. Stud. in Math., V. 11. Walter de Gruyter, 1989, xi + 225 pp, $69.95. [ISBN: 3-11-011734-7] Uses the Poincaré conformal model to reintroduce many of the classical results (proving them under weaker assumptions) of a subject which has become the key to understanding almost all 3-manifolds. JNC

**Algebraic Topology, S*(18), P*.** ***A History of Algebraic and Differential Topology, 1900-1960.*** Jean Dieudonné. Birkhauser Boston, 1989, xxi + 648 pp, $79. [ISBN: 0-8176-3388-X] An encyclopaedic description of the main events in the study of algebraic and differential topology (dimension $n > 4$) and their applications, including some detail on the work of the pioneers in the field. Divided into three parts: simplicial techniques and homology, the first applications of simplicial methods and of homology, homotopy and its relation to homology. Assumes familiarity with basic algebra and general topology. SB

**Algebraic Topology, P.** ***Lecture Notes in Mathematics-1375: Transformation Groups.*** Ed: K. Kawakubo. Springer-Verlag, 1989, viii + 394 pp, $33 (P). [ISBN: 0-387-51218-7] Proceedings of a conference held in Osaka in 1987 containing many papers on the algebraic and geometric aspects of groups acting on topological spaces. SG

**Topology, T(18: 1), S, P, L.** ***The Theory of Fixed Point Classes.*** Kiang Tsai-han. Springer-Verlag, 1989, xi + 174 pp, $69. [ISBN: 0-387-10819-X] Intended as "an easy and introductory treatise" on the topological theory of fixed points. Concerned largely with connected finite polyhedra. Based on J. Nielsen's work on fixed point classes. Appendices on homotopy, covering spaces, approximation theorems. Bibliography, epilogue. JS

**Operations Research, T(14: 1), L.** ***Mathematical Introduction to Linear Programming and Game Theory.*** Louis Brickman. Undergrad. Texts in Math. Springer-Verlag, 1989, xi + 130 pp, $34. [ISBN: 0-

387-96931-4] A mathematically rigorous exposition of the simplex algorithm, duality theory, and matrix games. JNC

**Optimization, P.** *An Atlas of Steiner Networks.* Sandra Lach Arlinghaus. Mono., No. 9. Institute of Mathematical Geography (2790 Briarcliff, Ann Arbor, MI 48105), 1989, v + 84 pp, $15.95 (P). Material drawn from author's unpublished Ph.D. thesis. Discusses the Steiner problem (find the shortest path joining a set of points), gives an algorithm which exhibits the detailed geometric construction of candidates for the Steiner network of the generalized Steiner problem, discusses the case for six points and enumerates the possible candidates for certain Steiner networks. LC

**Dynamical Systems, T(18), S, P.** *Practical Numerical Algorithms for Chaotic Systems.* Thomas S. Parker, Leon O. Chua. Springer-Verlag, 1989, xiv + 348 pp, $49.50. [ISBN: 0-387-96689-7] For those interested in nonlinear dynamics who lack background in chaotic dynamics or numerical techniques. Assumes familiarity with linear systems. Each chapter introduces a theoretical aspect of dynamical systems and then provides algorithms for computer-aided analysis of these systems. An appendix reviews concepts from differential topology used throughout the book. KS

**Dynamical Systems, T(16-17: 1), P, L.** *Elements of Differentiable Dynamics and Bifurcation Theory.* David Ruelle. Academic Pr, 1989, viii + 187 pp, $27.50. [ISBN: 0-12-601710-7] An introduction to differential dynamics. Emphasizes ideas rather than proofs with an eye toward applications to natural phenomena. Part one: basic definitions and results; Part two: bifurcation theory; Part three: 41-page appendix that includes basic background material gathered from analysis and topology, as well as an extension of the classical theory to the noncompact case. SP

**Dynamical Systems, P.** *Advanced Topics in the Theory of Dynamical Systems.* Ed: G. Fusco, M. Iannelli, L. Salvadori. Notes & Reports in Math. in Sci. & Eng., V. 6. Academic Pr, 1989, ix + 266 pp, $29.95. [ISBN: 0-12-269990-4] Proceedings of the conference held at Villa Madruzzo, Trento, Italy, June 1-6, 1987. SP

**Systems Theory, P.** *Lecture Notes in Control and Information Sciences-120: Large Scale Systems: Decentralization, Structure Constraints and Fixed Modes.* L. Trave, A. Titli, A. Tarras. Springer-Verlag, 1989, xiv + 384 pp, $49.70 (P). [ISBN: 0-387-50787-6] The title defines the topics; one needs to read no further than the Preface to see that English is not the first language of the writers, and no one has helped them with it. Their focus is on systems too large to permit all information to be accessible at a central control. AWR

**Stochastic Processes, S(18), P.** *Structured Stochastic Matrices of M/G/1 Type and Their Applications.* Marcel F. Neuts. Prob.: Pure & Appl., V. 5. Marcel Dekker, 1989, xiv + 510 pp, $125. [ISBN: 0-8247-8283-6] Presents a unified analysis and applications of the class of structured Markov chains and Markov renewal processes of $M/G/1$ type. Problem sets include engineering applications or extractions from articles on such. Some are suitable for masters' theses or class projects. Extensive bibliography. MLR

**Computer Science, P.** *Advances in Computers, Volume 28.* Ed: Marshall C. Yovits. Academic Pr, 1989, xi + 302 pp, $59.95. [ISBN: 0-12-012128-X] Five state-of-the-art expositions on design processes, fuzzy sets applied to artificial intelligence, parallel architecture applied to database systems, optical computing, and management intelligence systems. LAS

**Applications, S(15-16), P, L*.** *Popularizing Mathematical Methods in the People's Republic of China: Some Personal Experiences.* Hua Loo-Keng, Wang Yuan. Rev. & Ed.: J.G.C. Heijmans. Math. Model., No. 2. Birkhauser Boston, 1989, xxii + 209 pp, $49. [ISBN: 0-8176-3372-3] An elaboration of topics outlined in Hua's 1980 report to the International Congress on Mathematical Education giving examples of Hua's "popularization" lectures to managers and factory workers to promote efficiency through mathematical modelling. Examples include estimating mineral reserves, meshing gears, machine scheduling, and transportation problems. Includes Heine Halberstam's obituary of Hua, a preface by Hua, and an editor's preface. LAS

**Applications, P?** *Mathematical Geography and Global Art: The Mathematics of David Barr's "Four Corners Project."* Sandra L. Arlinghaus, John D. Nystuen. Mono., No. 1. Institute of Mathematical Geography (2790 Briarcliff, Ann Arbor, MI 48105), 1986, vii + 78 pp, $9.95 (P). Analysis and calculation required to locate precisely coordinates on the land mass of the earth of an inscribed tetrahedron with one vertex on Easter Island. These calculations formed the basis of a tetrahedral "sculpture" erected by David Barr with four precisely positioned tetrahedral vertices at the four corners of the inscribed tetrahedron. Includes generalizations to other platonic solids and refinements necessitated by variation in elevation and curvature of the earth's surface. LAS

**Applications, P*.** *The Collected Papers of Raymond D. Mindlin.* Ed: H. Deresiewicz, M.P. Bieniek, F.L. DiMaggio. Springer-Verlag, 1989, $130 set [ISBN: 0-387-96933-0]. *Volume I,* xxviii + 596 pp; *Volume II,* xxviii + 591 pp. At the time of his retirement in 1975, Raymond David Mindlin was James Kip Finch Professor of Applied Science at Columbia University. These two volumes contain, in order of appearance in print, photographically reproduced copies of over 100 of his research papers in applied mechanics published between 1934 and 1986. A seven-page bio/bibliographical sketch by his student and co-author, colleague and friend, H. Deresiewicz summarizes Mindlin's many contributions and honors. JDEK

**Applications (Biological Science), P.** *Lecture Notes in Biomathematics-79: Lindenmayer Systems,*

*Fractals, and Plants.* Przemyslaw Prusinkiewicz, James Hanan. Springer-Verlag, 1989, v + 120 pp, $19.80 (P). [ISBN: 0-387-97092-4] Exploration of capabilities of a special kind of parallel-process rewriting rules called *L*- (for Lindenmayer) systems which can explain the generation of such diverse creations as fractals, plants, formal languages, tilings, and DNA sequences. Focuses on distinctive role of parallel (rather than sequential) replacement. LAS

**Applications (Biological Science), S(15-16), L.** *Perspectives in Ecological Theory.* Ed: Jonathan Roughgarden, Robert M. May, Simon A. Levin. Princeton U Pr, 1989, vii + 394 pp, $22.50 (P); $60. [ISBN: 0-691-08508-0; 0-691-08507-2] Ecological theory addresses a wide range of topics, from the study of individuals through groups and populations of single species to entire ecosystems and biospheres. The editors of this collection of papers seek to create a comprehensive statement of what has been accomplished by ecology theory over its broad range of concerns. SM

**Applications (Biological Science), P, L*.** *Stochastic Processes in the Neurosciences.* Henry C. Tuckwell. CBMS-NSF Reg. Conf. Ser. in Appl. Math., V. 56. SIAM, 1989, v + 129 pp, (P). [ISBN: 0-89871-232-7] Lectures from a June 1986 conference held at North Carolina State University. Provides a complete introduction, beginning with reviews of relevant biology and deterministic models, then moving to classical stochastic models, and one-dimensional Markov diffusion processes. The final five chapters explain contemporary models, including stochastic partial differential equations, channel noise, Wiener kernel expansions, and neural networks (large populations of neurons). LAS

**Applications (Physical Science), P, L.** *Aperiodicity & Order, Volume 2: Introduction to the Mathematics of Quasicrystals.* Ed: Marko V. Jarić. Academic Pr, 1989, x + 226 pp, $59.50. [ISBN: 0-12-040602-0] Volume 2 of a multi-volume work intended as a "pedagogical introduction for beginners." Five chapters by seven contributors include an introduction to tiling by Majorie Senechal, a look at tilings and quasicrystals by R. Penrose, group theory of icosahedral crystals by P. Kramer and R.W. Haase, local properties of three-dimensional Penrose tilings by Andre Katz, and analysis of defects in quasicrystals and their kinetics by J. Bohsung and H.-R. Trebin. Numerous references. JDEK

**Applications (Physics), T(16-18: 1, 2), P, L.** *Solitons: An Introduction.* P.G. Drazin, R.S. Johnson. Texts in Appl. Math. Cambridge U Pr, 1989, xii + 226 pp, $59.50. [ISBN: 0-521-33389-X] This introductory text does not attempt to be completely rigorous, however, the numerous exercises and worked examples more than compensate. The prerequisites are modest: partial differential equations, calculus of variations, Fourier integrals, linear waves, Sturm-Liouville theory, and the hypergeometric function. MU

**Applications (Physics), S(18), P.** *Inverse Problems in Quantum Scattering Theory, Second Edition, Revised and Expanded.* K. Chadan, P.C. Sabatier. Springer-Verlag, 1989, xxxi + 499 pp, $79. [ISBN: 0-387-18731-6] In addition to updating the original material, two new chapters on spectral problems and numerical aspects have been added (*First Edition,* TR, August-September 1978). MU

**Applications (Physics), P.** *Variational Methods in Nonconservative Phenomena.* B.D. Vujanovic, S.E. Jones. Math. in Sci. & Eng., V. 182. Academic Pr, 1989, x + 371 pp, $59.95. [ISBN: 0-12-728450-8] Authors' objective is to dispel the notion that nonconservative mechanics has no Lagrangian structure. Authoritative exposition spiced with examples, historical comments, and references. JDEK

**Applications (Physics), P.** *Geometrical and Algebraic Aspects of Nonlinear Field Theory.* Eds: S. De Filippo, *et al.* Delta Series. North-Holland (US Distr: Elsevier Science), 1989, x + 248 pp, $79 (P). [ISBN: 0-444-87359-7] Proceedings of a conference held in Amalfi, Italy in May 1988. Algebraic geometry and topology are used to reformulate many problems and the physical theories in a unified setting. Contains 17 lectures. MLR

**Applications (Physics), T(16-17: 1), P, L.** *Rational Kinematics.* Jorge Angeles. Tracts in Nat. Philo., V. 34. Springer-Verlag, 1988, xii + 137 pp, $39.80. [ISBN: 0-387-96813-X] Written to draw upon college-level linear algebra and mechanics, this text quickly surveys much of the classical material from the elegant vantage point of invariant relations. The flavor is more of high-level algebra than physics, and astonishing results are common place. For example, after two pages of analysis of Coriolis acceleration, the author convincingly concludes that "contrary to a popular belief, it bears no physical significance." MU

**Applications (Simulation), T(17-18).** *Simulation Methodology for Statisticians, Operations Analysts, and Engineers, Volume I.* P.A.W. Lewis, E.J. Orav. Stat. & Prob. Ser. Wadsworth, 1989, xvi + 416 pp, $49.95. [ISBN: 0-534-09450-3] For majors in statistics, operations research, and engineering interested simulation techniques. A background in probability theory, stochastic processes, statistical theory, and computer programming recommended for some sections of the book. A software package for the IBM PC is available to execute the techniques discussed in the book. Examples and problems in both mathematical statistics and system simulation are provided. MSS

### Reviewers

RJA: Richard J. Allen, St. Olaf; SB: Steve Benson, St. Olaf; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; SG: Steven Galovich, Carleton; JDEK: Joseph D.E. Konhauser, Macalester; SM: Steve McKelvey, St. Olaf; JO: Jeff Ondich, St. Olaf; SP: Samuel Patterson, Carleton; MLR: Margaret L. Reese, St. Olaf; AWR: A. Wayne Roberts, Macalester; KS: Karen Saxe, St. Olaf; JS: John Schue, Macalester; LAS: Lynn Arthur Steen, St. Olaf; MSS: Myriam S. Steinback, Macalester; MU: Milton Ulmer, Carleton; PZ: Paul Zorn, St. Olaf.

# Modern Perspectives of Mathematics:

## Mathematics as a Consumer Good, Mathematics in Academia

## March 29-31, 1990, Cornell University

The Mathematical Sciences Institute (MSI), Cornell University, is bringing together influential mathematics consumers, educators, and researchers for a 3-day examination of the nation's use and development of mathematics resources. Drawn from government policy agencies, academia, and business and industrial laboratories, they will address how to develop mathematics education and research in ways that will encourage the interplay of mathematics and its applications in business, industry, and government.

### Lectures

**Malvin H. Kalos**, Cornell University, *The Impact of Computers on Mathematics,* (keynote address); **Edward E. David, Jr.**, EED, Inc., *Mathematics in a World of Mega Projects; Can it Survive?*; **Peter D. Lax**, Courant Institute of Mathematical Sciences, *Today's World and Mathematics;* **Jacques-Louis-Lions**, Centre National d'Etudes Spatiales College de France, *Pure and Applied Mathematics;* **Ivars Peterson**, Science News, *The Mathematical Tourist*; **Shmuel Winograd**, IBM, *Mathematics and Computers*; **Roger W. Brockett**, Harvard University, *Mathematics of Intelligent Machines*; **Mary F. Wheeler**, Rice University, *Relationships between Industry and Academics*; **Walter W. Hollis**, U.S.Army, *Mathematics in Defense -- Support to the Decision Process -- A Subset*; **John J. Hopfield**, California Institute of Technology, *Mathematics in Neural Sciences;* **Michael E. Fisher**, Univeristy of Maryland, *Mathematical Rigor in Physics: To What End?*

### Program Description

The symposium begins on Thursday, March 29, with registration, dinner, and the keynote address. Friday, March 30 will be devoted to lectures and discussion periods followed by an open, university lecture in the evening. The symposium will conclude with a panel discussion and a summary of recommendations on Saturday, March 31.

### Panelists

**Werner C. Rheinboldt** (Panel Chair), University of Pittsburgh; **George Metakides,** ESPRIT; **Cathleen S. Morawetz**, Courant Institute of Mathematical Sciences; **Malcolm R. O'Neill**, U.S. Army; **Richard S. Stepleman**, Exxon Research and Engineering Company; **Rose Teukolsky**, Ithaca High School; **Daniel Willard**, U.S. Army.

### General Information

The symposium will be held at Cornell University, Ithaca, NY. For information, contact conference secretary Diana Drake at the Mathematical Sciences Institute, Cornell University, 201 Caldwell Hall, Ithaca, NY 14853-2602, 607/255-7740.

### Organizing Committee

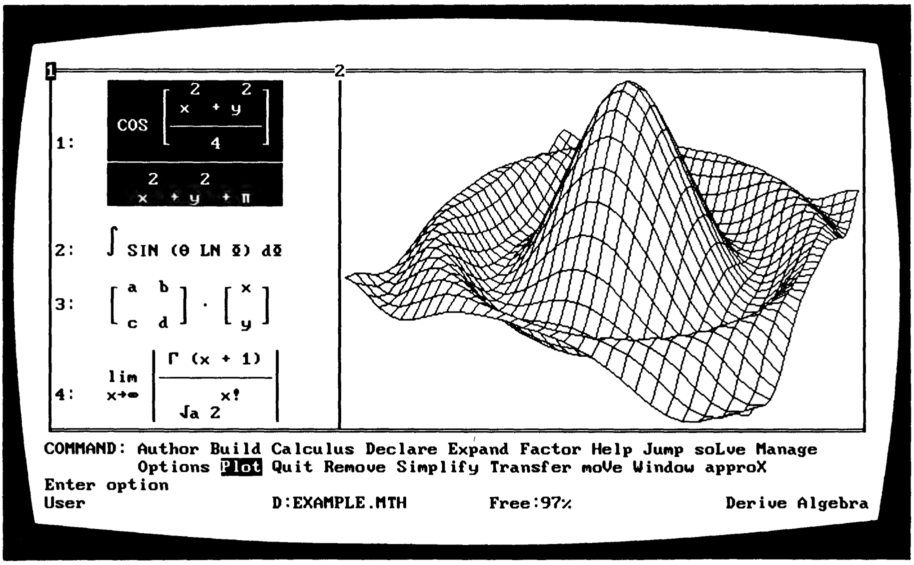
**Yervant Terzian** (Committee Chair), Chairman, Department of Astronomy; **Wilson V. Kone**, Associate Director for Administration, Mathematical Sciences Institute; **Simon A. Levin**, Director, Center for Environmental Research; **Anil Nerode**, Director, Mathematical Sciences Institute, **Lars B. Wahlbin,** Associate Director for Research, Mathematical Sciences Institute; all of Cornell University. **Daniel Willard**, U.S. Army.

MSI

MSI is partially funded by the U.S. Army Research Office. Cornell University is an affirmative action, equal opportunity educator and employer.

# DERIVE

## A Mathematical Assistant




*2000 years of mathematical knowledge on a disk*

***DERIVE*, the successor to muMATH, is a powerful computer algebra system for your PC compatible computer that provides the following capabilities:**

- **Exact and approximate arithmetic to thousands of digits**
- **Equations, complex numbers, trigonometry, calculus, vectors, and matrices**
- **2D and 3D function plotting with zooming capability**
- **MDA, CGA, EGA, VGA, and Hercules graphics and text support**
- **Attractive 2D mathematical display of formulas**
- **Easy to use menu-driven interface with on-line help**
- **Ideal for engineers, scientists, students and teachers**
- **$200 plus shipping: Call or write for information.**

**System requirements:** IBM PC or compatible computer, MS-DOS version 2.1 or later, 512K memory, and a 5¼ inch (360K) or a 3½ inch (760K) diskette drive. **Or** NEC PC-9801 or compatible computer, MS-DOS version 2.1 or later, 512K memory, and a 5¼ inch (640K) diskette drive.



**Soft Warehouse INC**

**3615 Harding Avenue, Suite 505 • Honolulu, HI 96816**
**(808) 734-5801 after noon PST**
***Handcrafted software for the mind.***

U S Postal Service

**STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION**
*Required by 39 U S C 3685)*

| 1A Title of Publication | 1B PUBLICATION NO | 2 Date of Filing |
|---|---|---|
| THE AMERICAN MATHEMATICAL MONTHLY | 0 0 0 2 9 8 9 0 | Sept. 29, 1989 |

| 3 Frequency of Issue | 3A No of Issues Published Annually | 3B Annual Subscription Pri |
|---|---|---|
| monthly except bi-monthly June/July and Aug/Sept | Ten | Library $90<br>Indiv. Member $2 |

4 Complete Mailing Address of Known Office of Publication *(Street City County, State and ZIP+4 Code) (Not printers)*
1529 Eighteenth St.,N.W., Washington, D.C. 20036

5 Complete Mailing Address of the Headquarters of General Business Offices of the Publisher *(Not printer)*
1529 Eighteenth St.,N.W., Washington, D.C. 20036

6 Full Names and Complete Mailing Address of Publisher Editor and Managing Editor *(This item MUST NOT be blank)*

Publisher *(Name and Complete Mailing Address)*
The Mathematical Association of America, 1529 Eighteenth St.,N.W., Washington, D.C. 200

Editor *(Name and Complete Mailing Address)*
Prof. Herbert S. Wilf, Univ. of Pennsylvania, Dept. of Mathematics, Philadelphia, PA 19

Managing Editor *(Name and Complete Mailing Address)*
Harry Waldman, MAA, 1529 Eighteenth St., N.W., Washington, D.C. 20036

7 Owner *(If owned by a corporation its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock If not owned by a corporation the names and addresses of the individual owners must be given If owned by a partnershi or other unincorporated firm its name and address as well as that of each individual must be given If the publication is published by a nonprofit organization its name and address must be stated ) (Item must be completed )*

| Full Name | Complete Mailing Address |
|---|---|
| The Mathematical Association of America, Inc. | 1529 Eighteenth St., N.W.<br>Washington, D.C. 20036 |

8 Known Bondholders Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds Mortgages or Othe Securities *(If there are none so state)*

| Full Name | Complete Mailing Address |
|---|---|
| None | |

9 For Completion by Nonprofit Organizations Authorized To Mail at Special Rates *(DMM Section 423 12 only)*
The purpose function and nonprofit status of this organization and the exempt status for Federal Income tax purposes *(Check one)*

(1) [X] Has Not Changed During Preceding 12 Months (2) [ ] Has Changed During Preceding 12 Months *(If changed publisher must submit explanation of change with this statement )*

| 10 Extent and Nature of Circulation *(See instructions on reverse side)* | Average No Copies Each Issue During Preceding 12 Months | Actual No Copies of Single Is Published Nearest to Filing Da |
|---|---|---|
| A Total No Copies *(Net Press Run)* | 22,815 | 22,738 |
| B Paid and/or Requested Circulation 1 Sales through dealers and carriers street vendors and counter sales | 0 | 0 |
| 2 Mail Subscription *(Paid and/or requested)* | 19,757 | 18,919 |
| C Total Paid and/or Requested Circulation *(Sum or 10B1 and 10B2)* | 19,757 | 18,919 |
| D Free Distribution by Mail, Carrier or Other Means Samples, Complimentary, and Other Free Copies | 716 | 817 |
| E Total Distribution *(Sum of C and D)* | 20,473 | 19,736 |
| F Copies Not Distributed 1 Office use, left over unaccounted spoiled after printing | 2,342 | 3,002 |
| 2 Return from News Agents | 0 | 0 |
| G TOTAL *(Sum of E F1 and 2—should equal net press run shown in A)* | 22,815 | 22,738 |

11 **I certify that the statements made by me above are correct and complete**

Signature and Title of Editor Publisher Business Manager or Owner
Kay Lamont
Membership Manager

PS Form 3526, Feb 1989 *(See instructions on reverse)*

"Light and witty without begin flippant, *The Calculus of Murder* is, with any luck, the first of a series." – *Liverpool Daily Post*

Dan Brodsky *is* back! The mathematician sleuth returns to attack an even deeper mystery in

# ADVANCED CALCULUS of MURDER

by

Erik Rosenthal
Department of Mathematics
University of New Haven

"*The Calculus of Murder* ... is the story of a mathematician who doubles as a ... private investigator. He does this outside work because he has not been able to latch onto a university teaching job. Perhaps his specialty – bounded linear operators on a separable Hilbert space – is too rarefied ... *The Calculus of Murder* turns out to be an orthodox and even classic mystery ... The writing is easy-going, the characteres well etched ..."

– *Newgate Callendar, New York Times Sunday Book Review*

In *Advanced Calculus of Murder*, Dan attends an operator theory conference at Oxford University in his role as research mathematician. However, he must don his private eye cap when the staid atmosphere is shattered by two brutal murders.

Published by ST. MARTIN'S PRESS at $18.95.

Special to readers of The *Monthly*: $9.95, postage included. A few copies of *The Calculus of Murder* remain and are specially priced at $11.95. (Foreign orders, please add $2.00.)

Send check or money order to: St. Martin's Press
P.O. Box 117
North Haven, CT 06473-0117

*Advanced Calculus of Murder* ☐ *The Calculus of Murder* ☐

Name ____________________

Address ____________________

City/State/Zip ____________________

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.